

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank) 2. REPORT DATE 6-30-95 3. REPORT TYPE AND DATES COVERED Final Tech. Report, 7/31/91 - 3/31/95

4. TITLE AND SUBTITLE A Collaborative Study of Turbulence-Generated Noise and Development of Fast Image Correlation Velocimetry Techniques 5. FUNDING NUMBERS N00014-91-J-1968

6. AUTHOR(S) Morteza Gharib, Principal Investigator

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Regents of the University of California
University of California, San Diego
Office of Contract & Grant Administration, 0934
La Jolla, CA 92093-0934 8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research
800 North Quincy Street
Arlington, VA 22217 10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited 12b. DISTRIBUTION CODE 19960822 241

13. ABSTRACT (Maximum 200 words)
The overall goal of this project has been to develop a new technique to predict hydrodynamic radiation from complex flows. The approach, which was first proposed in 1954 by H.W. Liepmann, relates the fluctuations at the edges of a flow field to its far-field acoustic signature. Only recently have time-resolved, global velocity measurement techniques been developed that allow investigation of Liepmann's method. In this project, we attempted to develop an understanding of various parameters for the model and develop new tools for their measurements. The approach we have taken has been to formulate the near-field displacement effects of shear flows such as jets and stratified shear layers, in the sense of boundary layer theory, and relate to the far-field sound pressure. To do this, we used laser-induced fluorescent imaging, digital particle image velocimetry, and scalar imaging velocimetry, to study temporal and spatial behavior of the various stratified and non-stratified flows.
Our accomplished tasks constitute major advances toward our stated goal:
1. Completed a major study of the dynamics of the stratified shear layers, especially as it relates to the most important scales for the Liepmann model.
2. Made extensive flow measurements of 3D submerged, temporal mixing layers.
3. Demonstrated the role of 3D vortical structures on the entrainment fluctuations.
4. Advanced the state-of-the art in image processing, specifically, the correlation of a pair of images, recorded in quick succession in time, to extract velocity-field information.
5. Designed and fabricated a dual-image CCD device to record a pair of images, as close as 1-2 microseconds apart, for the purpose of estimating the image in-plane velocity field in moderate to high-speed flows.

14. SUBJECT TERMS DTIC QUALITY INSPECTED 3 15. NUMBER OF PAGES 16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT 18. SECURITY CLASSIFICATION OF THIS PAGE 19. SECURITY CLASSIFICATION OF ABSTRACT 20. LIMITATION OF ABSTRACT

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

Final Report Summary

ARPA Contract #N00014-91-J-1968

The overall goal of this project has been to develop a new technique to predict hydrodynamic radiation from complex flows. The approach, which was first proposed in 1954 by H. W. Liepmann, relates the fluctuations at the edges of a flow field to its far-field acoustic signature. Only recently have time-resolved, global velocity measurement techniques been developed that allow investigation of Liepmann's method. In this project, we attempted to develop an understanding of various parameters for the model and develop new tools for their measurements.

The approach we have taken has been to formulate the near-field displacement effects of shear flows such as jets and stratified shear layers, in the sense of boundary layer theory, and relate to the far-field sound pressure. To do this, we used laser-induced fluorescent imaging, digital particle image velocimetry, and scalar imaging velocimetry, to study temporal and spatial behavior of the various stratified and non-stratified flows.

Our accomplished tasks constitute major advances toward our stated goal:

1. Completed a major study of the dynamics of the stratified shear layers, especially as it relates to the most important scales for the Liepmann model.
2. Made extensive flow measurements of 3D submerged, temporal mixing layers.
3. Demonstrated the role of 3D vortical structures on the entrainment fluctuations.
4. Advanced the state-of-the-art in image processing, specifically, the correlation of a pair of images, recorded in quick succession in time, to extract velocity-field information.
5. Designed and fabricated a dual-image CCD device to record a pair of images, as close as 1-2 microseconds apart, for the purpose of estimating the image in-plane velocity field in moderate to high-speed flows.

Final Technical Report

1. Structures in stratified mixing layers

In the first part of this investigation, a two-dimensional temporal mixing layer is generated with a stratified tilting tank similar to that used by Thorpe [J. Fluid Mech. **32**, 693 (1968)]. An extensive flow visualization study of the flow dynamics is made possible by matching the index of refraction of the two-layered fluid system by using Epsom salt and ethanol as the stratifying agents, and the two-dimensional density field is measured with the laser-induced fluorescence technique (LIF). It is found that the pairing process plays a crucial role in the mixing transition in a plane mixing layer. During the pairing period, the flow topography exhibits a highly complex behavior, with a elevated level of stretches and folds in the fluid interface within the region of the pairing vortical structures. A measurement of the material interface area between the two layers reveals an exponential growth. But as the pairing process becomes inhibited with increasing stratification, the flow topography grows less complex, and the mixing layer exhibits a less rapid linear interface growth. Also the ability of the mixing layer to mix the top and bottom fluid, as calculated from the measurements of the density field, is reduced with higher stratification. The reduced mixing is due in part to the reduction in the fluid entrainment into the Kelvin-Helmholtz vortices, the smaller entrainment ratio of the heavier bottom fluid into the cores, the reduced frequency of pairing of the large-scale structures, and the subsequent arrest of turbulence as the stabilization effect of the stratification eventually relaminarizes the flow. The stratified mixing layer exhibits many interesting instabilities, some of which are not observed in a homogeneous mixing layer--the baroclinically-generated secondary Kelvin-Helmholtz instability in the braid region of the main Kelvin-Helmholtz instability, gravitational convective instability of the Taylor-Rayleigh type within the large-scale vortices, vortex tearing, and vortex branching or dislocations of the Kelvin-Helmholtz vortices.

In the second part of this study, the effect of cross shear on a nominally two-dimensional plane mixing layer is investigated. A temporal mixing layer is generated with a stratified tilting tank; and cross shear, with the vorticity component in the streamwise direction, is introduced to this mixing layer to simulate situations that are widespread in industrial and natural flows. A new type of "co-rotating" streamwise vortices arises in the braid region of the mixing layer from the introduction of cross shear. The appearance and dynamics of these secondary co-rotating streamwise vortices are observed to be very similar to those of the primary Kelvin-Helmholtz instability, exhibiting similar roll-up and pairing, both qualitatively and quantitatively in the normalized length scale (wavelength) and time scale (time of pairing). Measurements of the density field by LIF show that mixing increases appreciably when a critical level of cross shear is introduced to the mixing layer, inducing the co-rotating streamwise vortices. The facts that the streamwise vortices reside in the braid and are roughly aligned along the principle axis of strain subject them to large vortex stretching, which means they are able to generate a considerable amount of turbulent stresses and small-scaled fluctuations. It is also found that if cross shear is introduced while the mixing layer is still essentially a plane two-dimensional shear layer, the mixing layer simply behaves as a normal two-dimensional mixing layer, with the flow redirected along an oblique axis. The co-rotating streamwise vortices and the corresponding increased mixing are produced only when cross shear is introduced around the time the primary shear layer starts to roll up into Kelvin-Helmholtz vortices and is no longer a plane two-dimensional shear layer. Evidence also shows that even in a two-dimensional stratified mixing layer without "global" cross shear, the co-rotating streamwise vortices can

develop where local cross shear is induced by streamwise baroclinic vorticity generation at locations with a high local curvature of the density interface and a large $\partial\rho/\partial z$ component. Two papers based on this work were submitted to the *Journal of Fluid Mechanics* (see attachments A & B).

2. Image Correlation Velocimetry

In this part of our work, we have undertaken a two-pronged effort to add to the technology of the use of a sequence of images of a quantity convected by a fluid to infer the flow velocity. In our work, under the sponsorship of the grant, we have

1. advanced the state-of-the-art in image processing, specifically, the correlation of a pair of images, recorded in quick succession in time, to extract velocity-field information, and
2. designed and fabricated a dual-image CCD device to record a pair of images, as close as 1-2 microseconds apart, for the purpose of estimating the image in-plane velocity field in moderate-to-high-speed flows.

The image-data-processing philosophy developed under this grant relies on the extraction of a displacement field that maps the first image onto the second. We have dubbed the method Image Correlation Velocimetry (ICV). In this approach, the ratio of the local (vector) displacement field to the time interval between the two images provides a (first-order accurate) estimate of the convecting velocity field. The method can be (and has also been) applied to images of particles, as employed in conventional DPIV techniques (e.g., Adrian 1991, Willert & Gharib 1991), but also to images of convected scalar fields, as, for example, arise in laser-induced fluorescence flow visualization images, typically, in liquid-phase flows, or Rayleigh, or resonant laser-scattering images in gas-phase flows. We expect to be relying on the latter in our future work in supersonic flows.

The development of the ICV methodology was initially performed by P. T. Tokumaru and P. E. Dimotakis. It is presently continuing, under AFOSR sponsorship, by G. G. Gornowicz and P. E. Dimotakis. A first publication documenting this part of the work has recently appeared in *Experiments in Fluids* and is included in this report as attachment C (Tokumaru & Dimotakis 1995).

As part of the second part of this effort, we have designed and fabricated a new, dual-image recording CCD device, in collaboration with J. R. Janesick, T. Elliot, and S. A. Collins of JPL. This device is based on CCD design and fabrication technology for a high-performance imaging device developed for the Cassini planetary mission and allows two images to be recorded as closely as a microsecond apart from each other. As of this writing, the devices were fabricated in a shared lot run at a silicon foundry (Reticon), have been tested on their wafers, to the extent possible, and are waiting packaging and further testing before being used in actual flow experiments for the first time. A brief mention of this was made in an IS&T/SPIE pre-publication, which is included as part of this report as attachment D (J. Janesick et al. 1995).

The Caltech team for this part of the effort was comprised of D. Lang and P. Dimotakis. The electronics, camera head, and digital image-data acquisition system for this high-speed camera have been developed in-house at Caltech and are nearing completion. Testing and a

first use of the new devices is expected in the next few months in an effort that continues under the same AFOSR sponsorship that is supporting the image-processing development effort.

References:

R. J. Adrian 1991 "Particle-imaging techniques for experimental fluid mechanics," *Ann. Rev. Fluid Mech.* vol. 23, 261-304.

J. Janesick, T. Elliot, R. Winzenread, J. Pinter, and R. Dyck 1995 "Sandbox CCDs," *IS&T/SPIE Symposium on Electronic Imaging: Science & Technology*, 4-10 February 1995 (San Jose, CA). Document #2415-01.

P. T. Tokumaru and P. E. Dimotakis 1995 "Image correlation velocimetry," *Exp. Fluids* vol. 19, 1-15.

C. E. Willert and M. Gharib 1991 "Digital particle image velocimetry," *Exp. Fluids* vol. 10, 181-193.

P. Atsavapranee and M. Gharib 1995 "Structures in stratified mixing layers Part. 1 The two-dimensional plane mixing layer," Caltech.

P. Atsavapranee and M. Gharib 1995 "Structures in stratified mixing layers Part. 2 The effects of cross shear," Caltech.

Structures in stratified mixing layers

Part 1. The two-dimensional plane mixing layer

By P. ATSAVAPRANEE AND M. GHARIB

Graduate Aeronautical Laboratories, California Institute of Technology,
Pasadena, CA 91125, USA

In the first part of this investigation, a two-dimensional temporal mixing layer is generated with a stratified tilting tank similar to that used by Thorpe [J. Fluid Mech. **32**, 693 (1968)]. An extensive flow visualization study of the flow dynamics is made possible by matching the index of refraction of the two-layered fluid system by using epsom salt and ethanol as the stratifying agents, and the two-dimensional density field is measured with the laser-induced fluorescence technique (LIF). It is found that the pairing process plays a crucial role in the mixing transition in a plane mixing layer. During the pairing period, the flow topography exhibits a highly complex behavior, with a elevated level of stretches and folds in the fluid interface within the region of the pairing vortical structures. A measurement of the material interface area between the two layers reveals an exponential growth. But as the pairing process becomes inhibited with increasing stratification, the flow topography grows less complex, and the mixing layer exhibits a less rapid linear interface growth. Also the ability of the mixing layer to mix the top and bottom fluid, as calculated from the measurements of the density field, is reduced with higher stratification. The reduced mixing is due in part to the reduction in the fluid entrainment into the Kelvin-Helmholtz vortices, the smaller entrainment ratio of the heavier bottom fluid into the cores, the reduced frequency of pairing of the large-scale structures, and the subsequent arrest of turbulence as the stabilization effect of the stratification eventually relaminarizes the flow. The stratified mixing layer exhibits many interesting instabilities, some of which are not observed in a homogeneous mixing layer--the baroclinically-generated secondary Kelvin-Helmholtz instability in the braid region of the main Kelvin-Helmholtz instability, gravitational convective instability of the Taylor-Rayleigh type within the large-scale vortices, vortex tearing, and vortex branching or dislocations of the Kelvin-Helmholtz vortices.

1. Introduction

The physics of a free plane mixing layer is of considerable interest from both fundamental and practical points of view. It has been widely suggested that mixing in the oceans and the atmosphere is partly governed by the turbulence generated by shear-induced instabilities of the Kelvin-Helmholtz type, as evident by the many field observations from various parts of the world. Woods (1968) observed shear-induced instabilities in the high shear and temperature gradient regions in the summer thermocline off Malta and concluded that the vertical heat flux in the thermocline is largely dominated by the presence of these instabilities, the attributes of which were found to be consistent with those of the Kelvin-Helmholtz instability in the theoretical study by Miles & Howard (1964). Browning & Watkins (1970),

through the use of a high power S-band radar, observed the evolution of Kelvin-Helmholtz instability from its inception to its turbulent breakdown in the strong shear region of a jet stream front, that among many other examples of atmospheric shear-induced instabilities presented in the comprehensive review by Fritts & Rastogi (1985). In engineering flows, mixing layers govern the aerodynamics of vehicles, the rate of mixing in chemical lasers and combustion chambers, and the spread of pollutants at industrial sites. Since the two-dimensional mixing layer is very generic, and its major features are shared by many other turbulent shear flows, the interest in its physics also extends to other shear flow research such as jets and wakes.

Studies of plane mixing layers date as far back as 1868 with the theoretical study by Helmholtz on the dynamical instability in a homogeneous shear flow with a step discontinuity in velocity. Kelvin later extended the work to shear flows with layers of different densities, hence the given name Kelvin-Helmholtz instabilities. But not until middle of this century was significant impetus in the field revived by the high quality detailed measurements of the state of the turbulent plane homogeneous mixing layer by Liepmann & Laufer (1947). Since then a great deal of research has been done on numerous aspects of the turbulent plane mixing layer. Various studies of flow dynamics, for low Reynolds number (Winant & Browand, 1974), high Reynolds number (Brown & Roshko, 1974; Dimotakis & Brown, 1976; and Koochesfahani & Dimotakis, 1986), and stratified (Koop & Browand, 1979) cases, have revolutionized our physical understanding by clearly demonstrating the roles of coherent large-scale vortical structures on the overall dynamics, the turbulent transition, and the turbulent transport processes.

Relatively recently, starting with the work of Konrad (1977) and subsequently with Breidenthal (1978, 1981), Bernal & Roshko (1986), and Lasheras & Choi (1988), the existence and the dynamics of three-dimensional coherent structures in a plane mixing layer have been studied in some detail with the recognition that they could have a crucial role in the turbulent processes. Konrad (1977) and Breidenthal (1978) found that the mixing transition is preceded by the appearance of spanwise sinuous disturbances of the Kelvin-Helmholtz vortices, downstream of which streamwise streaks are observed. Bernal & Roshko (1986) later determined that these streamwise streaks are streamwise counter-rotating vortices that are superimposed upon the primary, spanwise vortex structures, and they contribute not only to mixing of passive scalars but also to all components of the Reynolds stress. The flow visualization by Lasheras & Choi (1988) in a low Reynolds number shear layer with a sinusoidal spanwise forcing clearly serves as a strong supporting evidence of the earlier interpretations of the dynamics of the counter-rotating streamwise vortices.

Constant research in the field has also identified other possible mechanisms by which small-scale velocity fluctuations could be generated in turbulent mixing layers. In stratified mixing layers, gravitational instabilities, which give rise to convective rolls in the statically-unstable regions of the Kelvin-Helmholtz vortices where heavy fluid is raised over light fluid in the cat's eye spiral, is identified as an important mixing mechanism (Thorpe, 1971, 1985; Klassen & Peltier, 1985; and Schowalter, 1993). Browand & Troutt (1980) observed in a high-Reynolds-number turbulent mixing layer that local interactions between adjacent vortices can result in "branching" or "dislocation" of the vortex structures where one vortex structure can partly merge with another to form a tree-branch type pattern. The process of vortex tearing, in which a smaller Kelvin-Helmholtz structure is caught within the strain field between two larger vortices, have also been routinely predicted by many

researchers. Other secondary instabilities, such as the baroclinically-generated secondary Kelvin-Helmholtz-type instability (Thorpe, 1971, 1987; Staquet, 1994) and three-dimensional tubes and knots (Thorpe, 1987) can also contribute to the transport processes in a turbulent stratified mixing layer.

The current work is partly inspired by S.A. Thorpe's series of experiments (Thorpe, 1968, 1971, 1973,...) in which temporal mixing layers are developed from tilting an enclosed horizontal tube filled with a bottom layer of brine and a top layer of fresh water. Tilting the tube creates the potential for the two fluid layers to accelerate in opposite directions resulting in a velocity shear at the interface. In his papers, Thorpe describes in detail the dynamics of the two-dimensional Kelvin-Helmholtz instability that evolves from the shear layer, the process of pairing of the large-scale vortices, the eventual breakup into turbulence, and the relaminarization of the flow into a stable system of density layers in which the spreading rate levels off to zero rather than continuing at a constant rate as in the homogeneous case.

There are many advantages that Thorpe's experiments hold over the more common spatial mixing layer experiments. Thorpe's mixing layer is temporally-developing and thus lends itself better to comparison with numerical simulations and more closely resembles naturally-occurring mixing layers. There is no solid body disturbance at the interface such as those which a splitter plate would impose (wake component and surface roughness disturbances). And there is also no similar feedback effect from the end of the tunnel that would influence the flow dynamics at the splitter plate. However, experiments in a tilting tank are very tedious and time-consuming, prohibiting the researchers from doing certain type of studies, such as ones requiring large statistical sets of data. Nevertheless, the current work is carried out in a tilting tank, similar to that used by Thorpe.

The results of this study are presented in two parts. The first part of this series presents the experimental results on the basic two-dimensional plane mixing layer, and the second part focuses on a nominally-plane mixing layer that is complicated by the presence of an added cross shear. The main fundamental objective of the current work is to gain better understanding of the turbulent processes and how they are affected by stratification and cross shear. An established view is that for the flow to achieve random three-dimensional small-scale motions, it must experience a hierarchy of instabilities, both two-dimensional and three-dimensional *en route* to the turbulent transition. Much efforts in the field, therefore, go into identifying and describing these instabilities. Some specific questions that we hope to answer are:

- What are the roles of the different processes and instabilities involved in the turbulent transition?

- Besides the main instabilities (the primary K-H instability, the pairing instability, the Taylor-Rayleigh type convective instability, and three-dimensional streamwise instabilities), what is the significance, if any, of various secondary instabilities such as vortex dislocation, the secondary K-H instability, three-dimensional tubes and knots, and vortex tearing?

- What effects does stratification have on the entrainment, the scalar mixing and the turbulent state in the mixing layer?

- What are the effects on the mixing layer when cross shear is introduced (part 2)? Does the presence of cross shear alter the main features of the mixing layer or introduce any new instability?

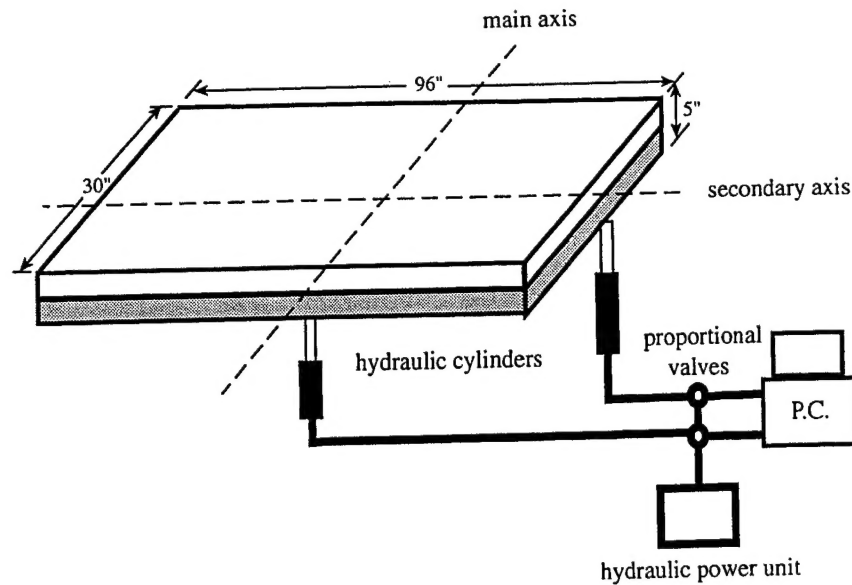


FIGURE 1. Tank schematic (not to scale)

2. Experimental apparatus and methods

2.1 The tilting tank

The experiment is performed in an apparatus similar to that used by Thorpe (1968, 1971, 1973, ...). The schematic of the setup is illustrated in figure 1. An enclosed tank, 96" long, 30" wide, and 5" deep is filled through four 1/2 inch-diameter openings on the bottom of the tank, first halfway with either fresh water or ethanol-water solution, and then slowly with epsom salt-water solution until full. In order to obtain a sharp interface, mixing between the two layers during the filling process is kept to a minimum with the use of a horizontal deflector plate placed slightly above the fill holes. At the interface a density gradient layer is formed whose thickness can be controlled by allowing the layer the appropriate time to grow with diffusion. The density difference between the two layers is varied between 0.005-0.08 g/cc, and the bottom layer takes about 1 hour to fill. The thickness of the density interface can be made as sharp as 4 mm. giving a spanwise aspect ratio of about 190. The stratified fluid is allowed to sit until the required density layer thickness is attained and all the disturbances from the filling process have died down, before the flow is generated by tilting the tank off the initially horizontal level.

Tilting of the tank provides the potential for the two layers to accelerate in opposite directions, resulting in a velocity shear at the interface. The tilting process is regulated by control signals from a PC. The control signals dictate the timing and the amount of opening of the proportional valves and the actuation of the hydraulic cylinders, providing the tank with smooth angular acceleration and deceleration in order to minimize vibrations and possible disturbances to the flow. The density stratification, density layer thickness, tilt angle, and the time the tank is left in the tilted position are chosen carefully to obtain the desired experimental parameters (Appendix 5.2). A certain time t_0 after the tank is tilted, when the desired experimental conditions (in terms of the Reynolds number and the Richardson number) are reached, the tank is tilted back to the horizontal position to create a steady flow for

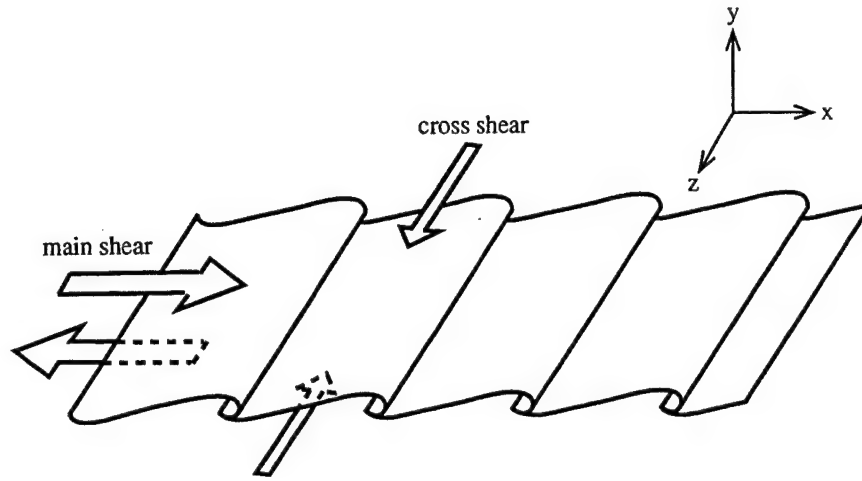


FIGURE 2. A schematic showing the overall picture of a mixing layer

the development of the Kelvin-Helmholtz instability which forms at the interface.

The tank is designed so that it can be tilted around two independent axes of rotations. For the study of the classic two-dimensional mixing layer, the base flow of the shear layer is generated by tilting the tank around only the main axis. To study the effect of cross shear on a nominally two-dimensional mixing layer, the tank is first tilted around the main axis, generating the main shear; and a moment later, with the primary shear layer still evolving, the tank is tilted around the secondary axis, for the cross shear. Variable degree of cross shear may be introduced precisely at any stage of the development of the main mixing layer, with the control of the PC. A schematic of the flow generated is represented in figure 2. The streamwise direction is defined as the direction of the main base flow (x-axis), and the spanwise direction is along the axis of the main Kelvin-Helmholtz vortices (z-axis).

The parameter ranges in most of the present experiments are such that the presence of density stratification has important dynamical consequences on the development of the mixing layer. The flow is characterized by the velocity difference of the mean flow and the density difference of the two layers. The shear layer has two vertical length scales: δ_v defined as the maximum velocity gradient thickness $\Delta U / (\partial u / \partial y)_{max}$ and δ_ρ , the maximum density gradient thickness $\Delta \rho / (\partial \rho / \partial y)_{max}$. δ_v is initially equal to δ_ρ right after the flow is generated by tilting the tank (because outside the layer with density gradient, the fluid would be of uniform velocity) but grows with diffusion as the free stream velocity accelerates (see Appendix 5.2). It generally does not increase more than 40% before the initial instability appears so that $1 \leq \delta_v / \delta_\rho \leq 1.40$. The flow is uniquely determined by these non-dimensional parameters:

$$Ri_i = g \Delta \rho \delta_v / \rho_{av} (\Delta U)^2; \quad Re_i = \Delta U \delta_v / \nu$$

$$Sc = \nu / D; \quad \delta_v / \delta_\rho$$

where ν and D are the kinematic viscosity and the molecular diffusivity of water, respectively. Ri_i and Re_i are the initial Richardson number and Reynolds number, measured at the onset of the Kelvin-Helmholtz instability, respectively. The Reynolds number is around 2000 for most experiments. The Richardson number ranges from 0.16 to 0.01. Sc is the Schmidt number which is fixed at about 600.

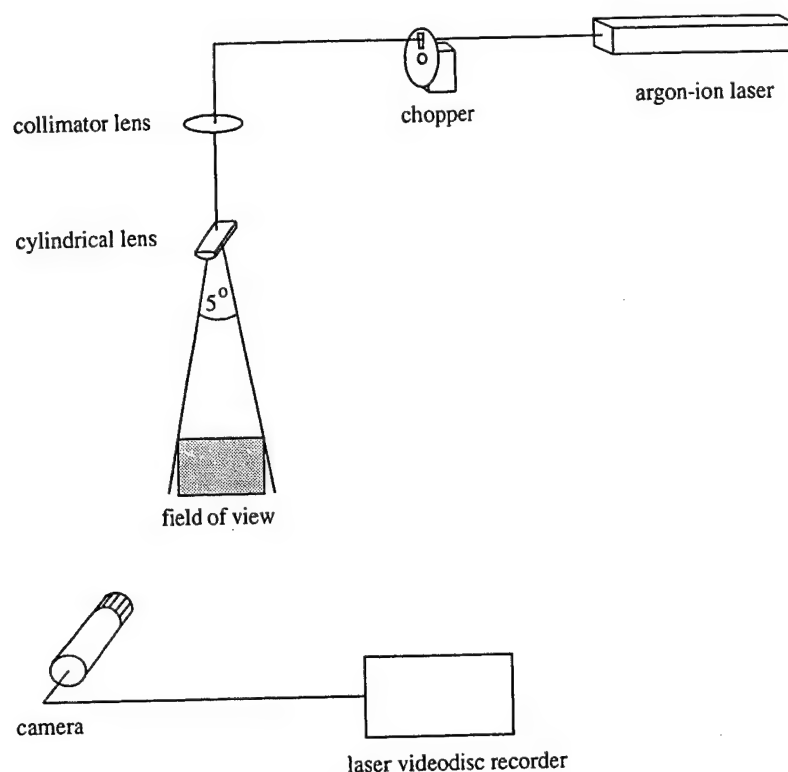


FIGURE 3. Flow visualization system

2.2 Index of refraction matching

For most of the experiments, ethanol-water solution is used as the top layer fluid instead of fresh water. It is necessary to match the index of refraction of the two layers for flow visualization and density-field measurement purposes; and as described by McDougall (1979), the use of two appropriate solutions can result in a match in the index of refraction between the two fluids while retaining the desired density stratification. McDougall found several appropriate choices of solutes, ethanol and table salt, epsom salt and sugar. In this study ethanol and epsom salt are also found to form an appropriate pair of solutes, yielding very good optical results and proving stable to double-diffusive instability of the salt-fingering type when the layers are stably-stratified. A typical density difference between the two layers is 0.02 g/cc, with the index of refraction of both layers being 1.3355. The solutions for the two layers are prepared by mixing the appropriate amount of solutes with water in separate tanks such that the desired density stratification is obtained and the indices of refraction of the two solutions are the same. A hand-held refractometer with an accuracy of 0.0001 is used to measure the index of refraction.

Care is also taken to insure that the temperature of the two premixed layers, measured with a high-resolution thermometer, matches to within 0.01 °C. This step eliminates any variation in the index of refraction due to thermal distortion or any convective instability arising from unstable temperature gradients. A peristaltic pump is used to fill the tank in order to minimize heating up the fluids. A simple fluid "capacitor" is designed and installed along the filling line so that the flow output experiences minimum pulsation from the peristaltic pump.

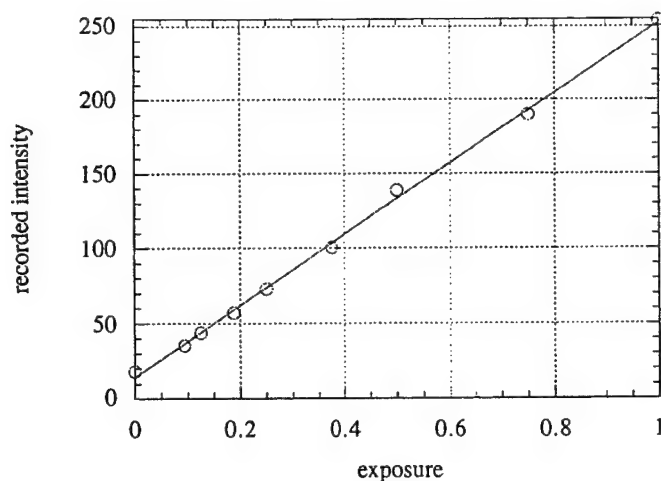


FIGURE 4. The response curve of the recording system

2.3 Laser-induced fluorescence (LIF)

Two-dimensional planes in the flow are visualized using laser-induced fluorescent dye. Sodium fluorescein (uranine) is premixed with the bottom layer fluid and its fluorescence is excited by an argon-ion laser sheet. The schematic of the flow visualization system is presented in figure 3. The beam of an argon-ion laser is collimated into a thin beam and passed through a cylindrical lens to form a sheet about 0.5-1 mm. in thickness. Special care is taken to achieve maximum uniformity of the laser sheet fluorescence by using a very small dye concentration of 4.2×10^{-7} M (molarity) in order to minimize laser attenuation due to dye absorption and long focal lengths for both the collimator lens and the cylindrical lens in order to minimize variations of the sheet intensity. Images are acquired with a Texas Instrument MC-1134GN CCD camera and recorded on a Sony laser videodisc recorder LVR-5000A at 30 frames per second. The response curve of the recording system, shown in figure 4, is found to be linear. The CCD array is 1134 pixels x 486 pixels, and the recorded images are digitized to 768 pixels x 480 pixels. The typical field of view of the images is 6.0 in. x 4.5 in., yielding a spatial resolution in the object plane of about 200 μm . x 240 μm . The laser beam is shuttered by a chopper, and the temporal resolution, which is the shutter speed, is 2 ms. Note that for dynamic conditions, the actual spatial resolution can be as much as 450 μm . x 300 μm . because of the smearing as fluid convects past the pixels within the 2 ms. exposure time.

2.4 Plan-view visualization

The plan view of the mixing layer is visualized in two ways. The first method is the standard shadowgraph technique, and the second method is similar to that used by Lasheras and Choi (1988). The interface between the two streams can be visualized by premixing the bottom layer with a high concentration of a fluorescent dye and illuminating from above with flash lamps to create an opaque interface. The concentration and light attenuation of the dye are so high that the illuminating light is not able to penetrate more than a few millimeters; therefore, the image obtained through this technique effectively visualizes the interface between the two layers, creating the appearance of a "muddy" bottom layer.

2.5 Density field and mixing measurements

The two-dimensional plane LIF images obtained as described in section 2.3 are used to calculate the two-dimensional density field of the mixing layer, in a way similar to the line density measurements by Koochesfahani & Dimotakis (1986). As described in section 2.3, special care is taken to achieve maximum uniformity of the laser sheet and minimum laser dye attenuation. If the illuminating laser sheet is perfectly uniform in the observation plane and the laser dye attenuation is infinitesimal, the density field can be easily obtained since the recorded intensity on each pixel would be directly proportional to the fluorescence from the corresponding sampling volume with a linear recording system. In turn the fluorescence from each sampling volume is directly proportional to the fractional volume of the dyed fluid at that location. It follows then that the density field can be directly obtained from the recorded intensity field. The local volume fraction of the bottom fluid is given by

$$\xi = v_b / (v_t + v_b) = (I - I_t) / (I_b - I_t)$$

where I is the intensity recorded from the sampling volume, I_t is the noise level recorded from a sampling volume with no dye (top fluid), and I_b is the fluorescence intensity of the unmixed dyed fluid. The density of the fluid in the sampling volume is, therefore, given by

$$\rho = \rho_b (1 + \xi \Delta \rho)$$

In order to utilize the two-dimensional LIF images in the calculation of the density field, the issues of the laser sheet non-uniformity and the laser dye attenuation need to be addressed. The long focal lengths of the collimator lens and the cylindrical lens result in a very nearly parallel laser sheet, with a maximum fan angle of only 2.5° from vertical. Therefore, the laser sheet can be roughly estimated by a parallel laser sheet which has a horizontal Gaussian profile in intensity as given by,

$$I(x, Y_o) = I_{max}(Y_o) \cdot [1 - f(x)]$$

The measured dropoff function $f(x)$ is kept small with a maximum less than 0.1. In the vertical direction the local laser sheet intensity in the y -direction as attenuated by the fluorescent dye is given by

$$I_{absorp}(X_o, y) = I_{no\ absorp}(X_o) \cdot \exp\left(-\varepsilon_o \int_0^y C_d dy\right)$$

where ε_o is the dye molar absorption coefficient, and C_d is the local molar dye concentration.

Due to the small dye concentration used, the maximum attenuation is only about 6%. A test image is taken before each run to obtain the dye molar absorption coefficient and the horizontal Gaussian profile. The intensity of each pixel is then corrected for the Gaussian profile dropoff and the laser dye attenuation as followed:

$$I_{corrected}(x, y) = I_{measured}(x, y) / \left\{ [1 - f(x)] \cdot \exp\left(-\varepsilon_o \int_0^y C_d dy\right) \right\}.$$

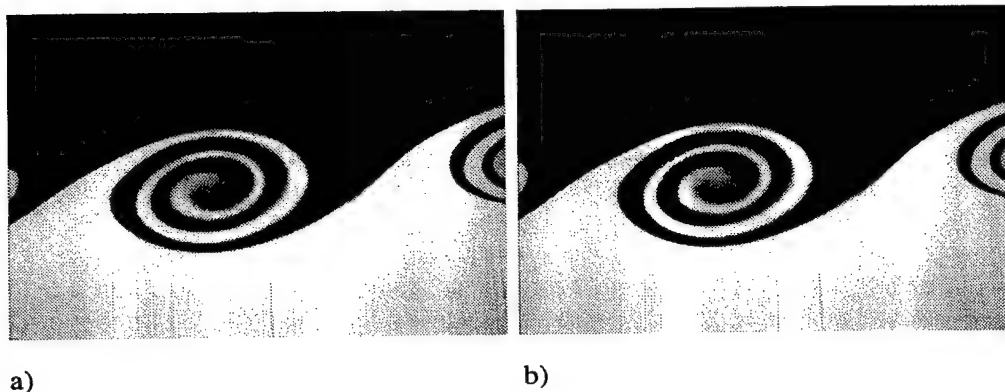


FIGURE 5. (a) non-corrected image (b) image corrected for horizontal Gaussian laser sheet profile and laser sheet attenuation by dye absorption

The method works remarkably well if the laser sheet is approximately parallel and the errors that need to be corrected are small to begin with. Figure 5(a) shows an example of a raw image, and figure 5(b) shows the corresponding corrected image, both false-color coded for easy detection of intensity variation.

2.6 The effect of resolution

A fundamental problem associated with the current image recording technique is that the information on the eddies with scales smaller than the spatial resolution is lost. The Kolmogorov scale, for current experimental conditions, is on the order of $50 \mu\text{m}$. The Batchelor scale, the smallest diffusive scale, in a two-dimensional mixing layer is estimated (Corcos & Sherman, 1984) to be $\lambda_d = (\pi D/2\gamma)^{1/2}$, which, under current experimental conditions, is about $20 \mu\text{m}$. Therefore, the measurement resolution, about $450 \mu\text{m} \times 300 \mu\text{m}$, is severely limited and cannot resolve the smallest mixing scale when the flow turbulence is active. This difficulty, however, does not exist for mixing measurements done by recording the density field a few minutes after the mixing layer has gone through the turbulent transition, and the flow has relaminarized. By this time, the fluid velocity fluctuations have died down, and all the sharp gradients have been smeared out. This measurement represents the "overall" mixing attained in the mixing layer and is analogous to those made by Koop & Browand (1979) in a spatial stratified mixing layer in which the mixing is measured with a conductivity probe at the farthest downstream station where the flow has relaminarized and attained its "asymptotic" amount of mixing.

3. Results and discussions

3.1 Rollup and pairing

A general discussion of the mixing layers as observed through flow visualizations is presented in this section and section 3.2. The general characteristics of the developing Kelvin-Helmholtz instability are mainly consistent with the findings of past investigations. The flow manifests the usual dynamics seen in spatial stratified mixing layer experiments: the growth of the most-amplified wavelength of the disturbances and the distortion into nonlinearity, the rolling up of the vorticity

interface into Kelvin-Helmholtz vortices, the pairing of the large-scale structures, the turbulent transition, and the eventual relaminarization of the flow into a stable system of density layers. The density thickness of the diffusive interface is obtained before each experiment by calculating an equivalent mixedness thickness δ_m , defined like a displacement thickness, from an LIF image as described in Appendix 5.1. The density layer thickness δ_ρ is roughly twice that of δ_m (and would be exactly twice for a piecewise-linear density profile). The averaged δ_ρ calculated in this way is 0.60 cm. and agrees well with the value measured by traversing a conductivity probe. The dimensionless wavenumber α is found to be independent of the Richardson number, with an average value of 0.40 and a standard deviation of 0.05.

A few seconds after the tank is tilted, a uniform array of Kelvin-Helmholtz waves is observed. Shown in figures 6, 7, and 8 are planar cross sections of mixing layers of similar Reynolds numbers (around 2000) and Richardson numbers of 0.012, 0.022, and 0.043, respectively. The instability waves grow with time and eventually become nonlinear with the initially uniform layer of vorticity rolling up into cat-eye structures. At these Richardson numbers the initial rollup of the Kelvin-Helmholtz instability seems qualitatively unaffected by the stratification, up to the point of the maximum rollup or the "climax" state. Note the almost symmetric nature of the Kelvin-Helmholtz vortices in figure 6. Roughly equal amount of heavy and light fluids are entrained into the large-scale structures, as opposed to a spatial mixing layer where the faster fluid is more abundant in the cores.

As the Richardson number is further increased, the appearance of the large-scale Kelvin-Helmholtz vortices becomes significantly altered. In general, as the stratification effect increases, a markedly-reduced amount of the heavier bottom fluid is entrained into the cores; and the maximum size of the Kelvin-Helmholtz vortices at the climax state is reduced. Figure 9 illustrates the appearance of the large-scale vortices at the climax state for two different Richardson numbers, 0.043 and 0.079. The ratio of the bottom fluid to the top fluid in the core, varies from about 1 for $Ri=0.012$ to 0.8 for $Ri=0.043$ (figure 9a) and 0.5 for $Ri=0.079$ (figure 9b). The outer band of the core material in vortex to the left of figure 9b cannot even complete the cat's eye, further highlighting the influence a moderate stratification can have on the dynamics of the Kelvin-Helmholtz vortices. The maximum roller height to wavelength ratio is measured at different Richardson numbers, with the Reynolds number constant at about 2000 and plotted in figure 10, along with the data from Thorpe (1973). The data show that as the mixing layer becomes more stratified, the Kelvin-Helmholtz vortices entrain less fluid into the smaller cores, which are the turbulent and mixing regions of the mixing layer. Consequently it is implied that the resultant amount of mixing would also be reduced, as is in fact shown in the mixing measurements presented in section 3.4. Beyond the Richardson number of around 0.16 the Kelvin-Helmholtz instability is totally suppressed, in line with the findings of Koop & Browand (1979).

Beyond the climax state, the effect of stratification on the dynamics of the large-scale vortical structures becomes evident in another manner. The average number of pairings that each Kelvin-Helmholtz vortex experiences becomes significantly reduced with higher stratification. The vortices in the $Ri=0.043$ case experience about 0.7 pairing each on average (number of vortical structures involved in pairings divided by the total number of structures), while the vortices in the $Ri=0.012$ case always experience a pairing, about 1.2 pairing each. When compared with the data by Koop & Browand (1979), the large-scale vortices in our experiments on average

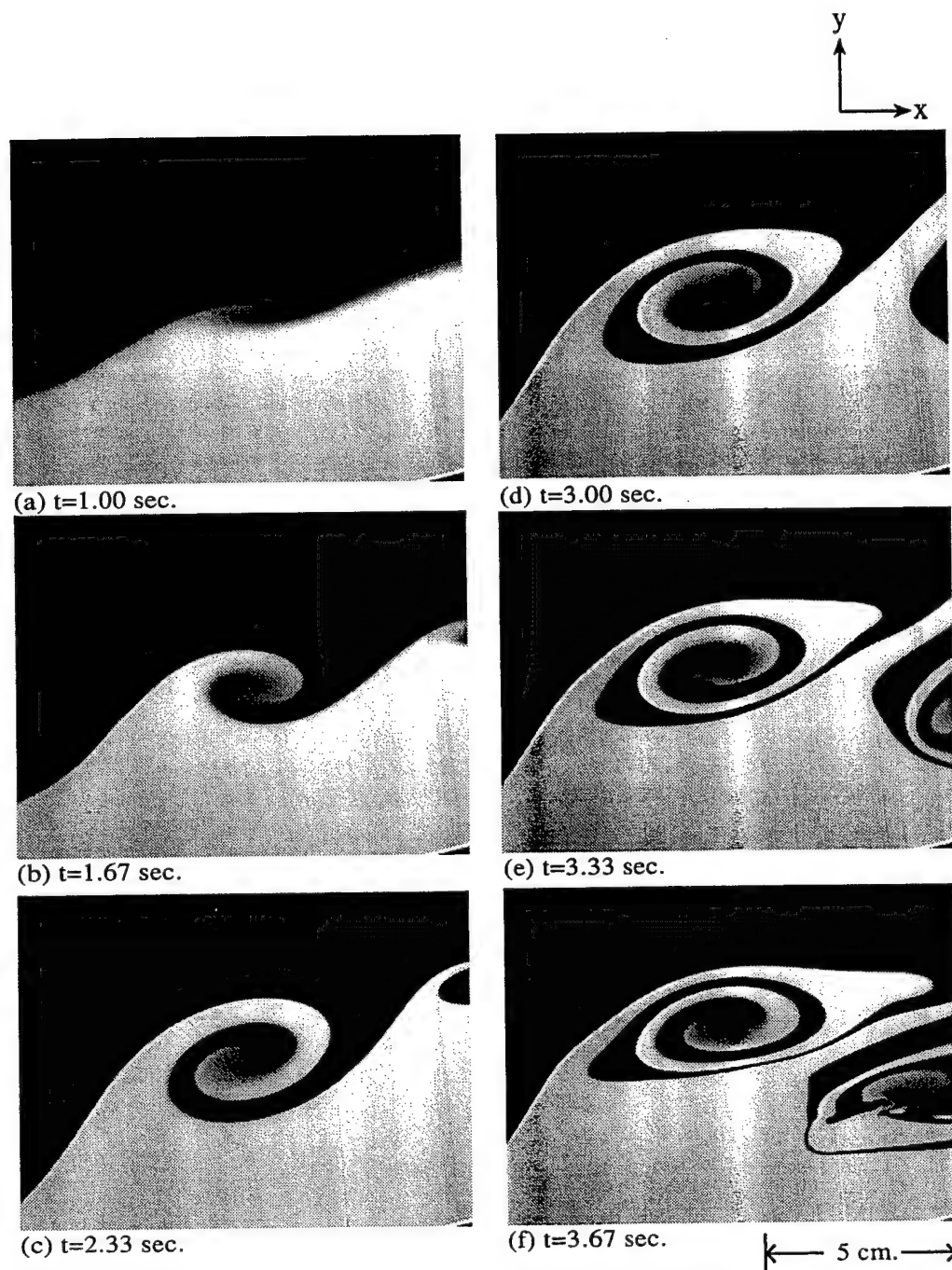
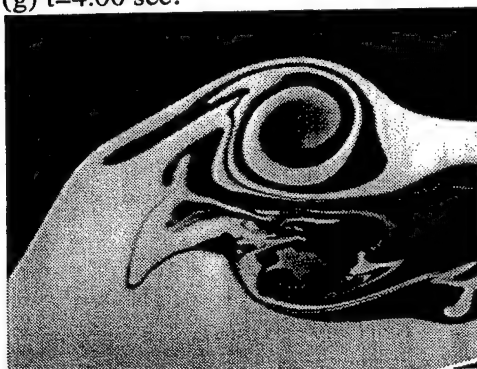
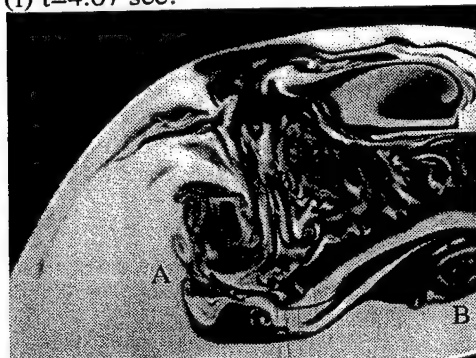
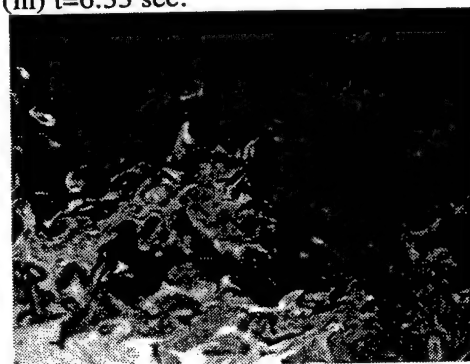


FIGURE 6. Side view showing the rollup, pairing, and breakdown of the Kelvin-Helmholtz vortices in a weakly-stratified two-dimensional mixing layer with $Re=2150$ and $Ri=0.012$. The time label for each image is from the onset of the Kelvin-Helmholtz instability. Scale is at the bottom right of the image.

(g) $t=4.00$ sec.(k) $t=5.33$ sec.(h) $t=4.33$ sec.(l) $t=5.67$ sec.(i) $t=4.67$ sec.(m) $t=6.33$ sec.(j) $t=5.00$ sec.(n) $t=8.00$ sec.

← 5 cm. →

Figure 6

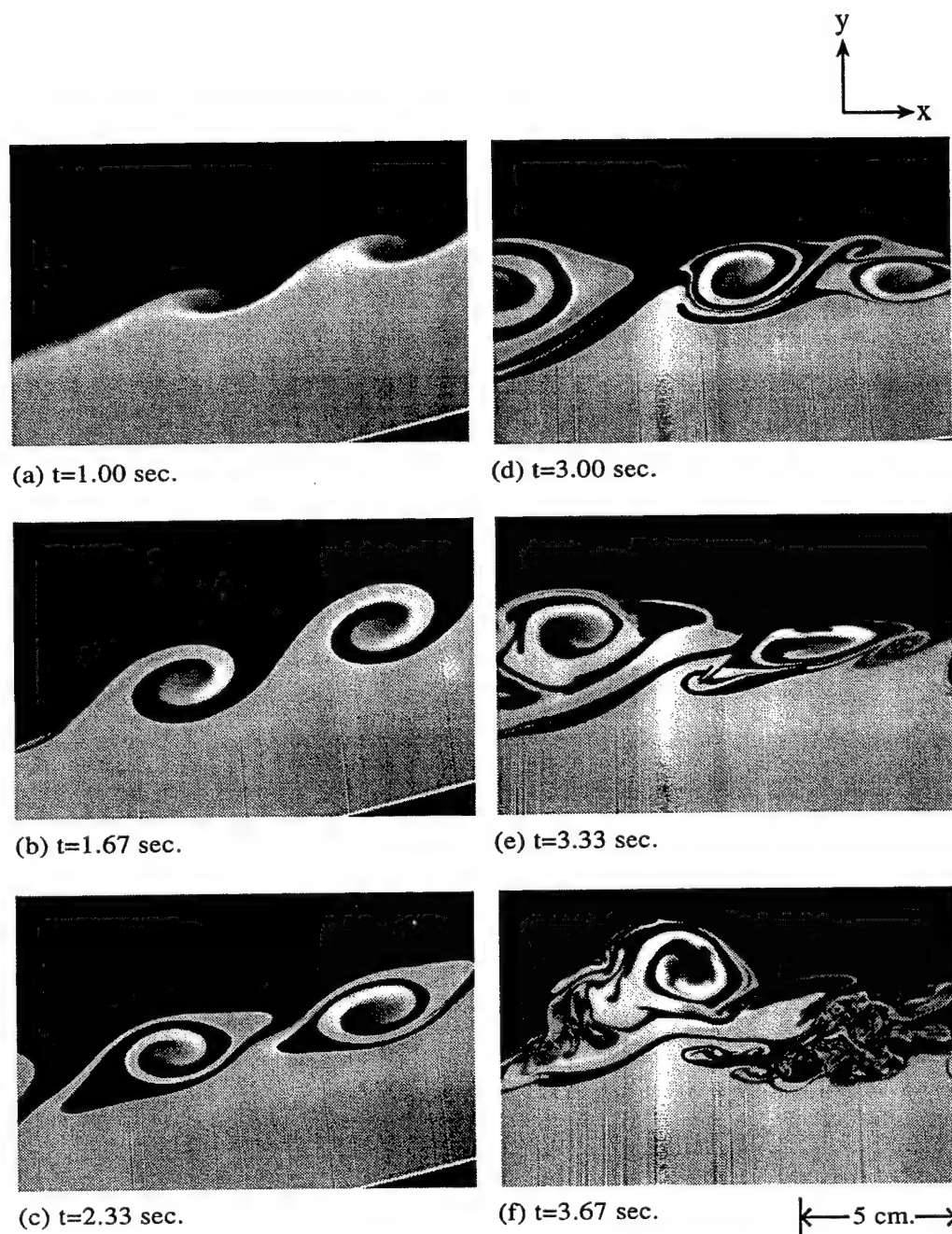
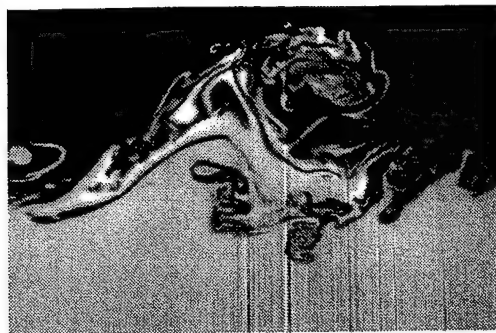
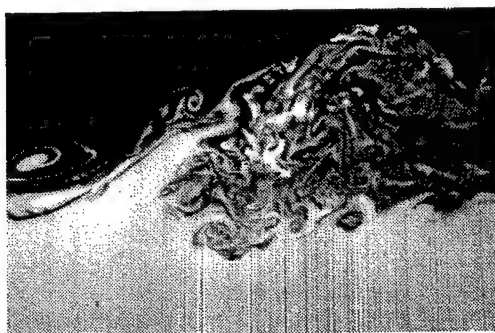
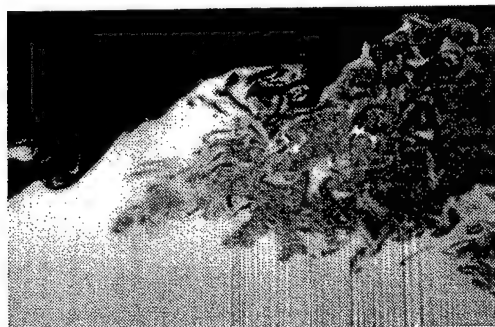
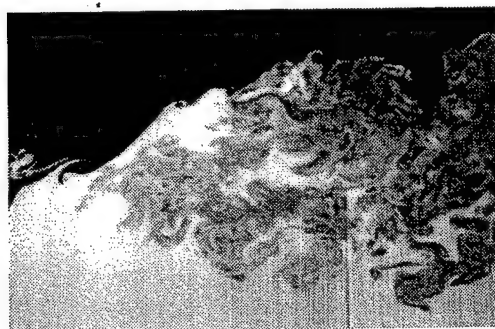


FIGURE 7. Side view showing the rollup, pairing, and breakdown of the Kelvin-Helmholtz vortices in a stratified two-dimensional mixing layer with $Re=2020$ and $Ri=0.022$. The time label for each image is from the onset of the Kelvin-Helmholtz instability. Scale is at the bottom right of the image.

(g) $t=4.00$ sec.(k) $t=5.33$ sec.(h) $t=4.33$ sec.(l) $t=5.67$ sec.(i) $t=4.67$ sec.(m) $t=6.33$ sec.(j) $t=5.00$ sec.(n) $t=8.00$ sec.

← 5 cm. →

Figure 7

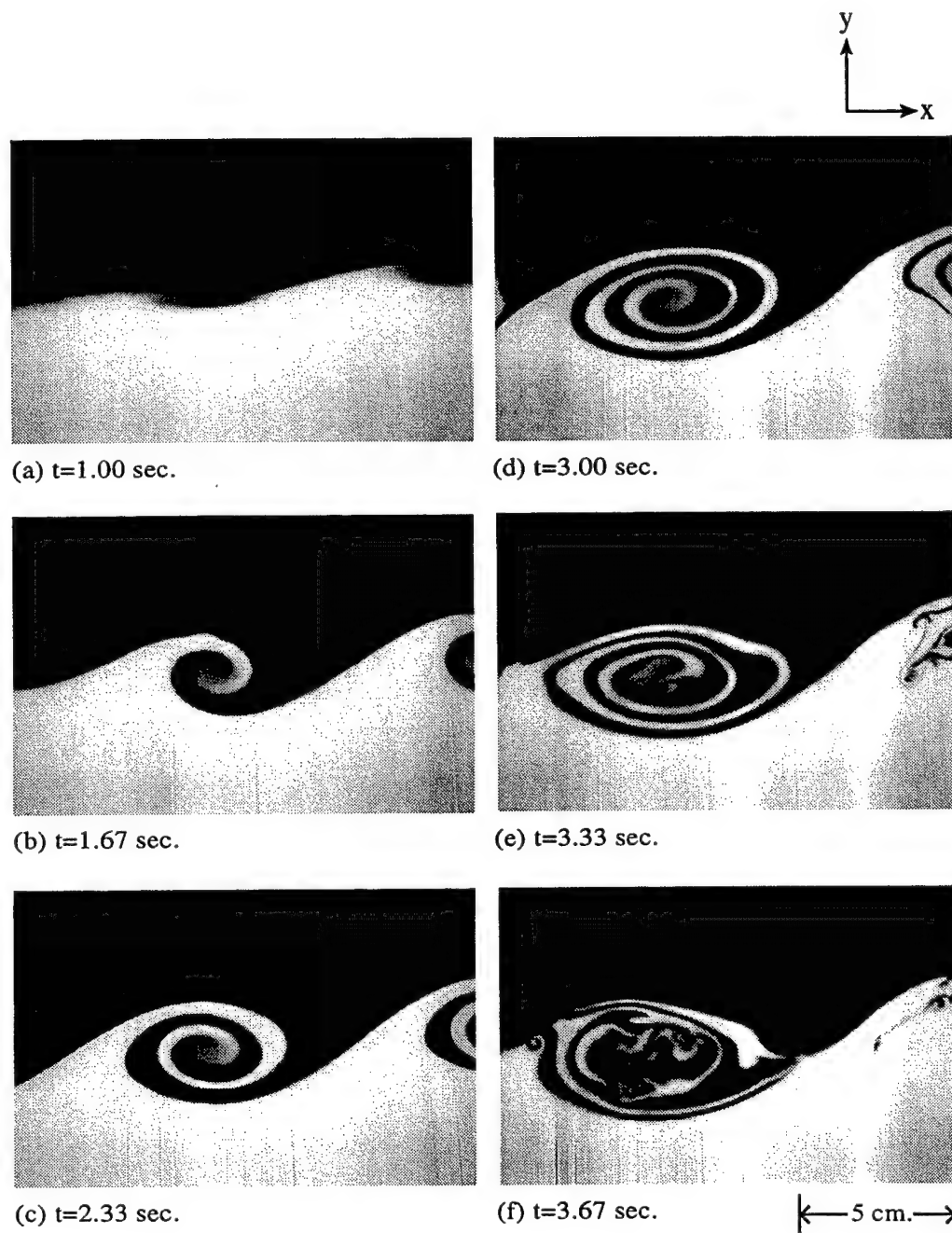
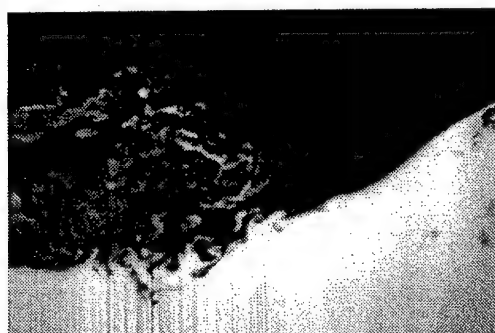
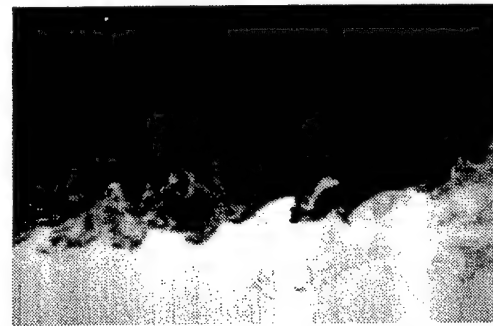
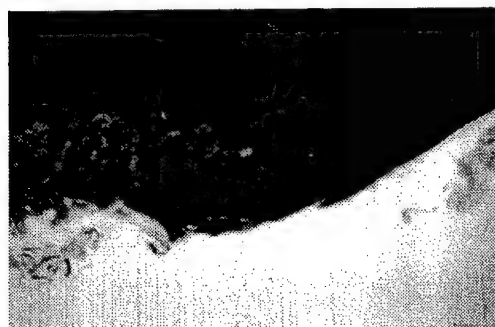


FIGURE 8. Side view showing the rollup, pairing, and breakdown of the Kelvin-Helmholtz vortices in a stratified two-dimensional mixing layer with $Re=2010$ and $Ri=0.043$. The time label for each image is from the onset of the Kelvin-Helmholtz instability. Scale is at the bottom right of the image.

(g) $t=4.00$ sec.(k) $t=5.33$ sec.(h) $t=4.33$ sec.(l) $t=5.67$ sec.(i) $t=4.67$ sec.(m) $t=6.33$ sec.(j) $t=5.00$ sec.(n) $t=8.00$ sec.

← 5 cm. →

Figure 8

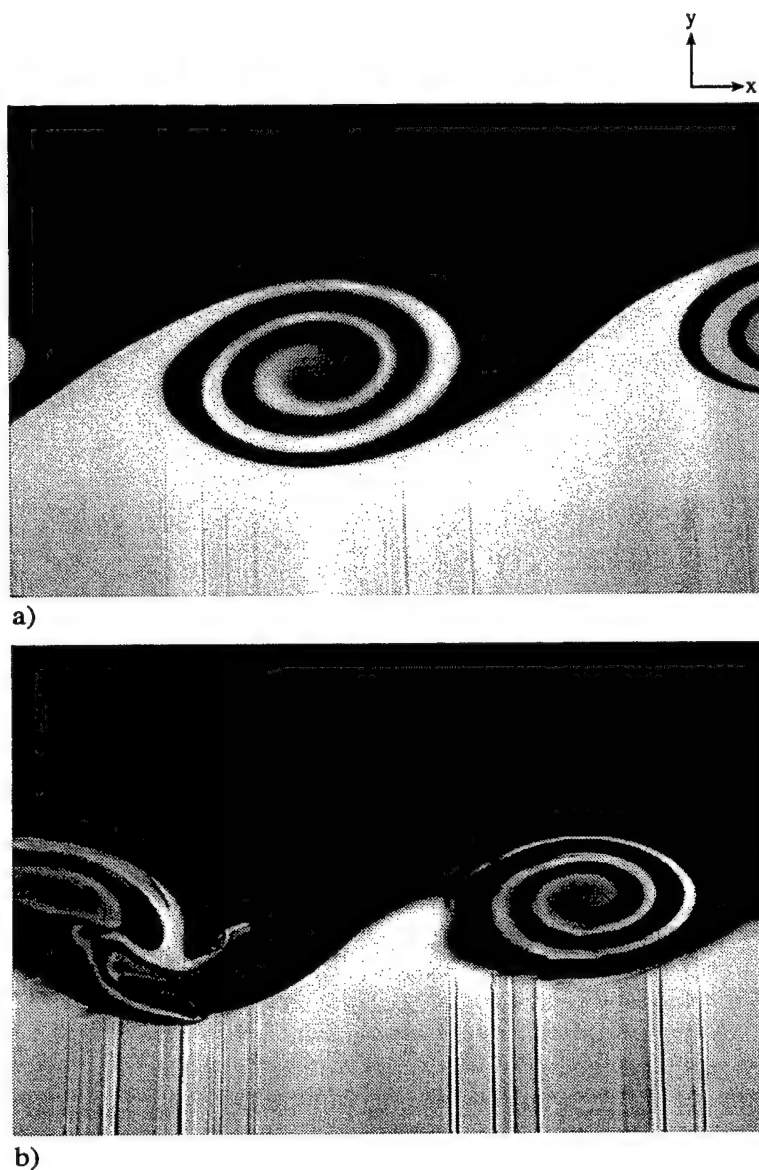


FIGURE 9. Side view of the Kelvin-Helmholtz vortices at the climax state for two Richardson numbers. (a) $Re=2010$ and $Ri=0.043$ (b) $Re=1900$ and $Ri=0.079$

experience about one fewer pairing than those in the spatial mixing layer experiments at similar Richardson number but much lower Reynolds number of 300. This is possibly due to two effects: the higher Reynolds number in our experiments and the boundary effect of the tank. At a higher Reynolds number the vortices make the turbulent transition after fewer pairings (in the limit of infinite Reynolds number, a spatial mixing layer will go turbulent right away at the tip of the splitter plate, without any pairing). Once the vortices have gone through the turbulent transition, they will transfer much of their energy to small scales, leaving less rotational energy to induce pairings of the large-scale vortices. Also the top and bottom boundaries of the tank are at a distance such that, according to analysis by Hazel (1972), any pairing after

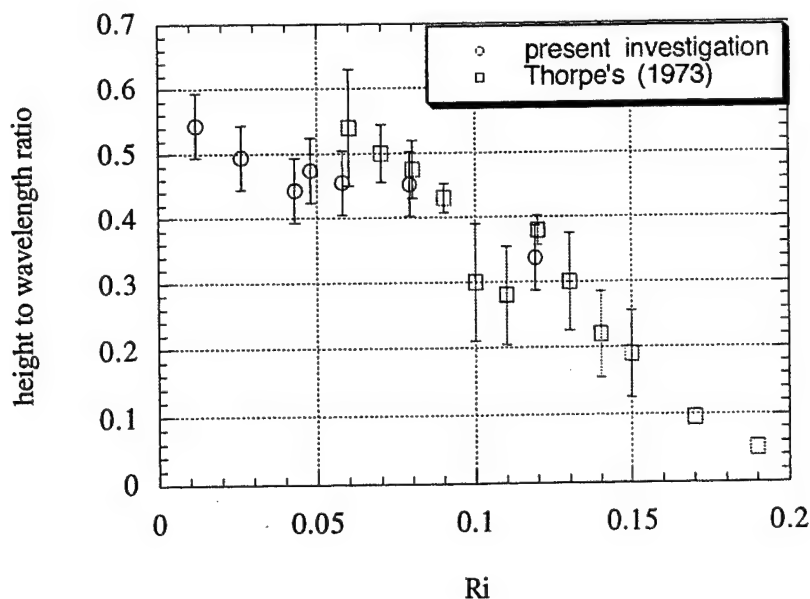


FIGURE 10. The variation of the maximum Kelvin-Helmholtz roller height (at climax state) to wavelength ratio with Richardson number.

the first would be dampened. The initial growth of the shear layer and the first pairing should be largely unaffected.

The most important consequence of stratification on the mixing layer is perhaps the stabilizing effect on the growth of the Kelvin-Helmholtz instability and its subsequent dynamics. The available kinetic energy drawn from the mean flow to amplify the initial disturbances is now partly converted into potential energy in the task of uplifting the heavy bottom fluid and depressing the light top fluid in the rollup and pairing processes. At higher stratification, the large-scale structures entrain less fluid into the cores and start losing their ability to interact with their neighbors. Also beyond the turbulent transition, it is observed that the stratified mixing layer does not follow a linear growth like its homogeneous counterpart. Consequent arrest of turbulence, relaminarization and restratification of the flow follow as the turbulent flow is eventually dampened by the stabilizing effect of stratification. Beyond this point, there is little further growth in the mixing layer.

3.2 Secondary shear instabilities, convective instability, vortex dislocation, tubes & knots, and vortex tearing

Within the braid region (the area between two Kelvin-Helmholtz vortices) of the mixing layer, an interesting type of shear-induced secondary instability commonly develops. As observed in figures 6m, 7j, 8l, and 11, there can be anywhere from one to as many as five or six such small Kelvin-Helmholtz-like vortical structures between two primary vortices. Except for the smaller size, these vortical structures are very similar to the main Kelvin-Helmholtz structures in appearance and dynamics, exhibiting similar rollup and pairing dynamics. Several previous works have provided evidence for the existence of this type of secondary instability. Thorpe (1968) has reported observing this type of instability in his tilting tank experiments and depicts its appearance in a drawing in his 1987 paper. Many other investigators,

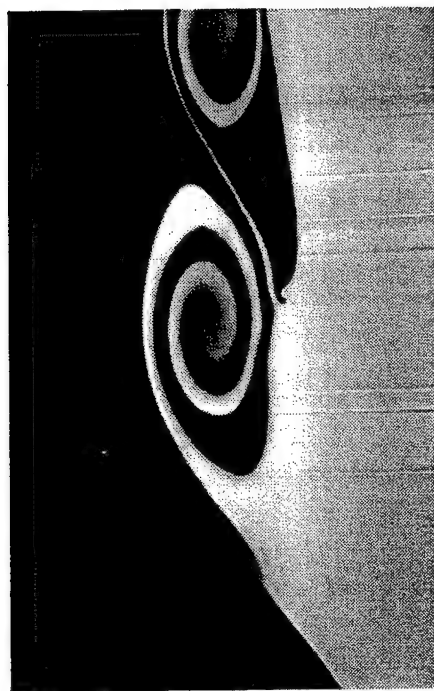
such as Patnaik *et al.* (1976), have also speculated on the possibility of its existence in stratified mixing layers, and the three-dimensional numerical simulations of stratified shear layers by Staquet & Riley (1989) have also identified the existence of these secondary spanwise Kelvin-Helmholtz modes.

The presence of this type of secondary shear instability in a stratified mixing layer and its absence in a homogeneous one is clearly due to the difference in the way vorticity is generated. The vorticity generation in a stratified mixing layer, unlike the homogeneous case, is baroclinic and given by the source term $-(g/\rho_0)\partial\rho/\partial x$. Downstream of a splitter plate in a homogeneous mixing layer, there is no new generation of vorticity, only advection and diffusion. In a stratified mixing layer, however, the baroclinic vorticity generation continues on through the rollup process and has a local maximum around the stagnation point in the braid (where $\partial\rho/\partial x$ is at a local maximum). Along the braid, the local shear induced by this mechanism, coupled with the thinly-stretched interface induced by the strain field of the main Kelvin-Helmholtz vortices, can result in a low local Richardson number which allows for a Kelvin-Helmholtz-type instability. For example the local Richardson number around the braid in figure 8i, just before the appearance of the secondary braid instability, is estimated to be around 0.06. The small scale of the secondary instability is consistent with the local scale of the interface on which it develops. Note in figure 11d that the secondary instability can repeat itself, creating even smaller tertiary structures (arrowed) within the braid region between a pair of secondary vortices. This behavior raises a very interesting question of whether this cascade can continue as the Reynolds number of the mixing layer increases.

Other secondary shear instabilities associated with high local shear are also observed within the cores of the primary structures. Note the two Kelvin-Helmholtz-like vortex structures that form around one of the pairing large-scale structures, marked 'A' in figure 6j. In the next frame 6k, those two vortices have merged into one in a pairing process. Another vortex structures of somewhat larger scale is observed at 'B.' There is no evidence that these vortex structures are results of baroclinic vorticity generation, and it is quite possible that the same type of structures can also be active in homogeneous mixing layers.

A moderate amount of stratification can also result in a gravitational instability of the Taylor-Rayleigh type (Thorpe, 1971, 1985; Klassen & Peltier, 1985; and Schowalter, 1993). This type of instability has been reported to be in the form of convective rolls within the statically-unstable regions of the Kelvin-Helmholtz cores. In the current experiments, convective instabilities have been observed in both the two-dimensional convective rolls (as seen in the plan-view of the shadowgraph in figure 12f, marked 'C') and the three-dimensional finger-type forms (as seen in figure 8f and 11b). These gravitational instabilities result from a system of multiple layers of alternately-stacked heavy and light fluids. At the layer where heavy fluid is found on top of light fluid, convective instabilities may amplify with the effect of gravity. Although the existence of stratification is usually thought of as having a dampening effect on the small-scale processes, this is one instance in which it can have a contribution by allowing the stored potential energy in the statically-unstable regions to be converted into small-scale kinetic energy through the amplification of gravitational instabilities.

A plan view of the interface between the top and bottom layers (figure 13), reveals a very striking feature. Although the flow is generated in a very low-disturbance environment, the resulting spanwise rollers are often time not perfectly two-dimen-



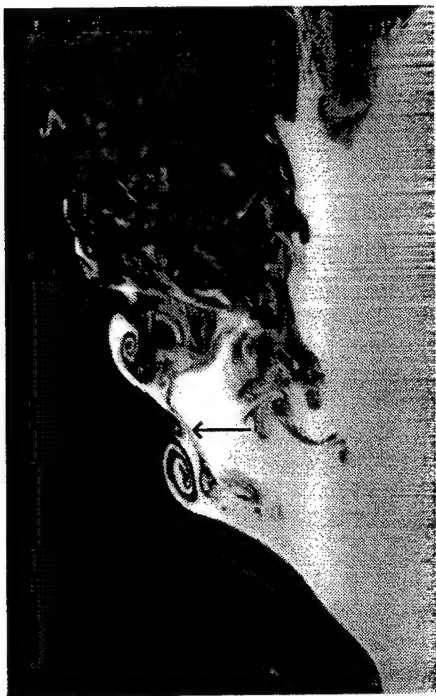
a) $t=3.67$ sec.



b) $t=4.67$ sec.



c) $t=6.10$ sec.



d) $t=6.43$ sec.

5 cm.

FIGURE 11. Side view showing secondary Kelvin-Helmholtz vortices in a stratified mixing layer with $Re=2200$ and $Ri=0.026$

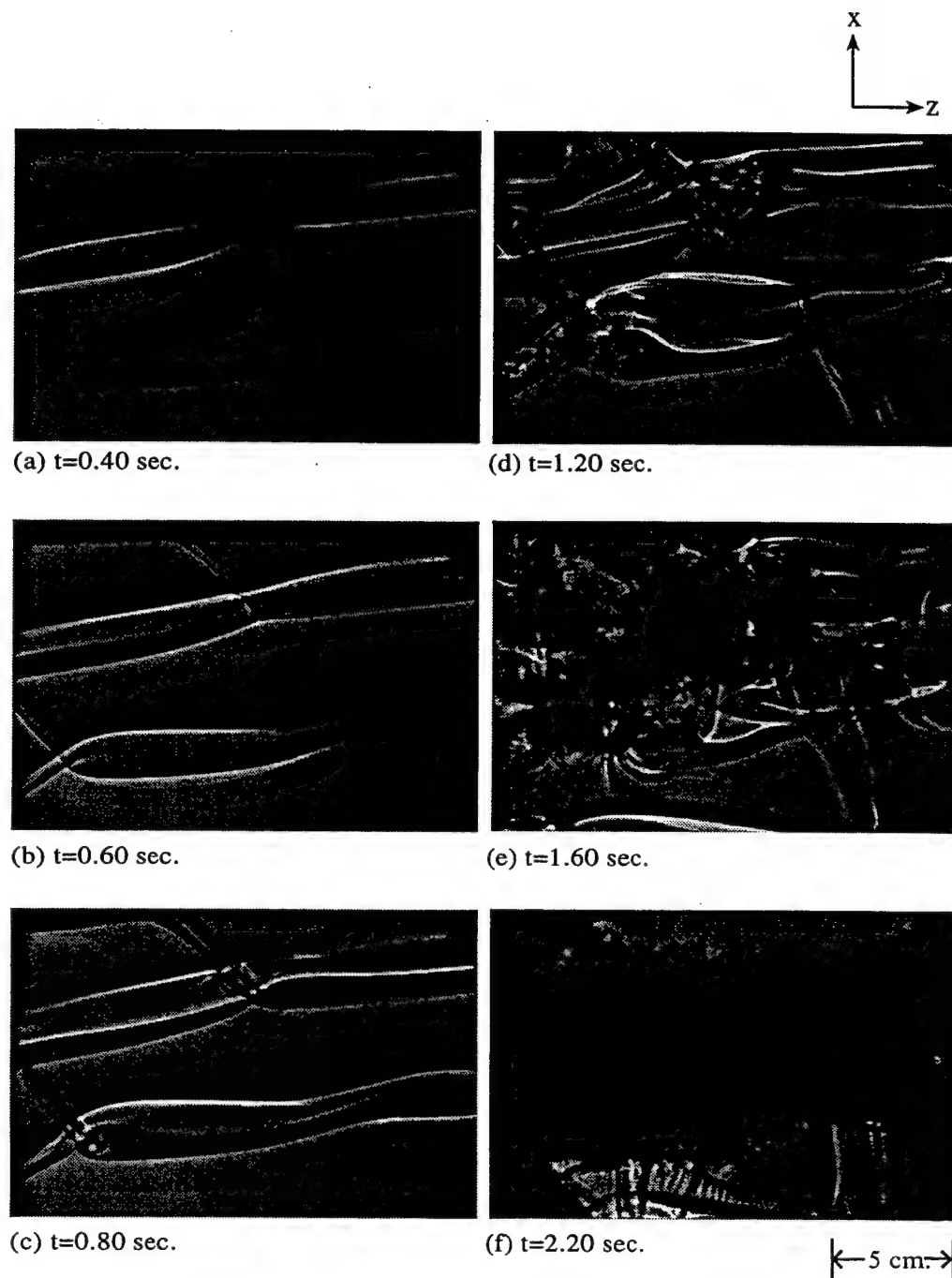


FIGURE 12. Plan-view shadowgraph of a two-dimensional stratified mixing layer with $Re=2010$ and $Ri=0.043$

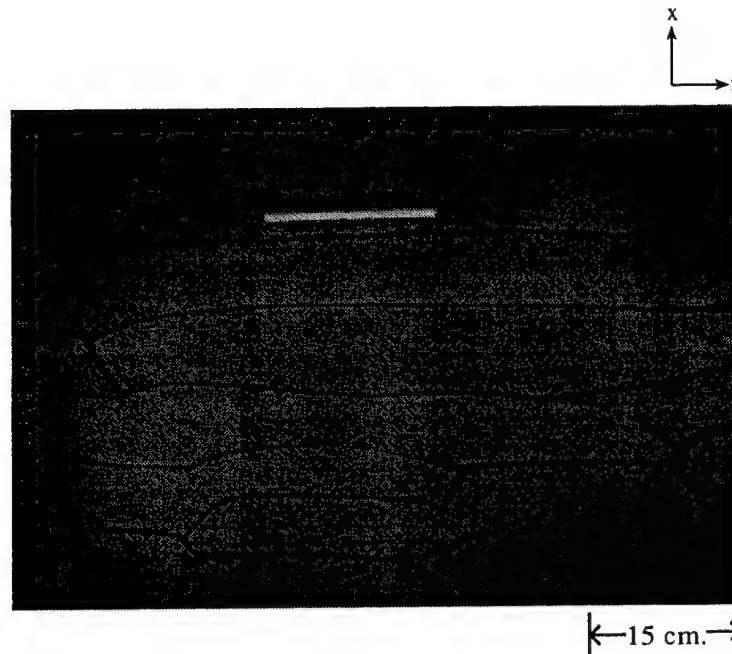


FIGURE 13: Plan view showing the "branching" or "dislocation" patterns in the initial rollup of the Kelvin-Helmholtz instability. The bottom layer is premixed with a high concentration of dye so that the illumination does not penetrate the interface more than a few millimeters. The black lines represent the front edge of the K-H rollers that would form shadow when illuminated from top. Reynolds number for the mixing layer is 2800, while the Richardson number is 0.040.

sional. The picture of the flow is reminiscent of patterns in desert sand dunes: the structures in general are spanwise but exhibit random defects or dislocations of the Kelvin-Helmholtz vortices, similar to the behavior observed in a high Reynolds number homogeneous mixing layer by Browand & Troutt (1980). It is observed that the Kelvin-Helmholtz structures, in the early stages of rollup, tend to develop into separate interacting "patches" in the plane of the shear layer. This patch behavior is clearly observed in the plan view of the shadowgraph in figure 12. In the early stages, the visible length of the vortices usually does not span the whole width of the tank indicating the existence of regions in which the amplitude of the vortices is larger in the middle than around the edges. The phase and the wavelength of the vortices in a patch sometimes slightly mismatch those in the neighboring patches, resulting in vortices that appear staggered. This deviation from perfect periodicity in the Kelvin-Helmholtz vortices has been attributed to the presence of a "phase jitter," which arises from the amplification of subharmonic perturbations from the background noise [Ho, *et. al.* (1991)]. Since the background noise is random in the shear plane, its amplification could conceivably result in slightly different vortex evolution patterns in different regions of the mixing layer. Consequently, dislocations form in the interacting regions as the vortices try to negotiate the phase and wavelength mismatches.

Perfect examples of dislocations are illustrated in figure 12 (marked 'D' and 'E'). Thorpe (1985) has described similarly-looking structures in some detail which he calls 'knots'. Perhaps, it is not clear if they can be categorized as different types of

structures. The 'tubes' structures (marked 'F') are also routinely observed by Thorpe (1985). The most important observation about these structures, whether they are categorized as vortex dislocations, tubes, or knots; however, is that small-scale transition is always observed to initiate at these locations.

Figure 14 shows an interesting situation in which a small kelvin-Helmholtz vortex is caught within the strain field of two larger vortices. The high strain rate of the larger vortices cause the smaller vortex to stretch along the principle axis and flatten along the perpendicular direction, causing it to collapse. This type of "vortex tearing" is rarely observed and probably does not contribute significantly to the overall turbulent processes in the mixing layer, although it is possible that this phenomenon happens more frequently at smaller scales which has not been recorded through flow visualization.

3.3 Material interface stretching measurements

The side-view image sequences in figures 6-8 show the dynamics of two-dimensional mixing layers with different levels of stratification. The most striking contrast among the three cases is that the observed scalar field for the case of $Ri=0.043$ is far less complicated than those of $Ri=0.012$ and $Ri=0.022$. This distinction is a direct consequence of the fact that the large-scale vortices in the $Ri=0.043$ case did not pair. The pairing process in the two lower Ri cases is observed to be much more than merely solid-body rotation of two large vortices around each other. As the two vortices are brought closer together in their interaction, each one stretches in the strain field of the other, resulting in thinly-stretched interface which elongate in one direction while contracting accordingly in the perpendicular direction. The stretched interface then folds upon itself creating a high level of convoluted surfaces. This complicated stretching and folding of the material interface between the two layers, as seen in figure 6 g-i and figure 7 e-g, does not resemble any known dynamical flow instability and seems to be mainly a kinematic phenomenon. In fact the appearance of the flow patterns resembles patterns in flows that are considered chaotic, such as those in the numerical calculations of two-dimensional mixing layers by Jimenez & Martel, 1991 (figure 10) and Min, 1994 (figures 5.5 and 5.13). One feature of a chaotic flow is the presence of Smale horseshoe functions (Chien *et al.*, 1986). The Smale horseshoe functions involve the stretching and folding of a square onto itself, and similar behavior in the material line shapes can be seen in figure 6 h-j. Another feature of a chaotic flow is that the stretching of a material line is exponential, with the implication that chaotic flows 'stir' the fluids better than non-chaotic flows by generating a greater amount of diffusive interface.

The measurement of the growth of the material interface between the top and the bottom layers is carried out by digitizing each LIF image, thresholding it to the intensity that represents the fluid of the density equal to $(\rho_{bottom} + \rho_{top})/2$, finding the edges of the thresholded images, and measuring the total length of the interfacial lines which is proportional to the total diffusive surface area. A sample of the resulted images is presented in figure 15 which is to be recognized as the same experiment of figure 6. The result of the material interface measurement is presented in figure 16. The abscissa represents the time in seconds normalized with the shear layer characteristic time $\lambda/(\Delta U/2)$. The ordinate represents the material-interface area, where the two fluid layers are in contact, normalized by the initial value before the Kelvin-Helmholtz rollup. Strong interactions of the two pairing vortices result

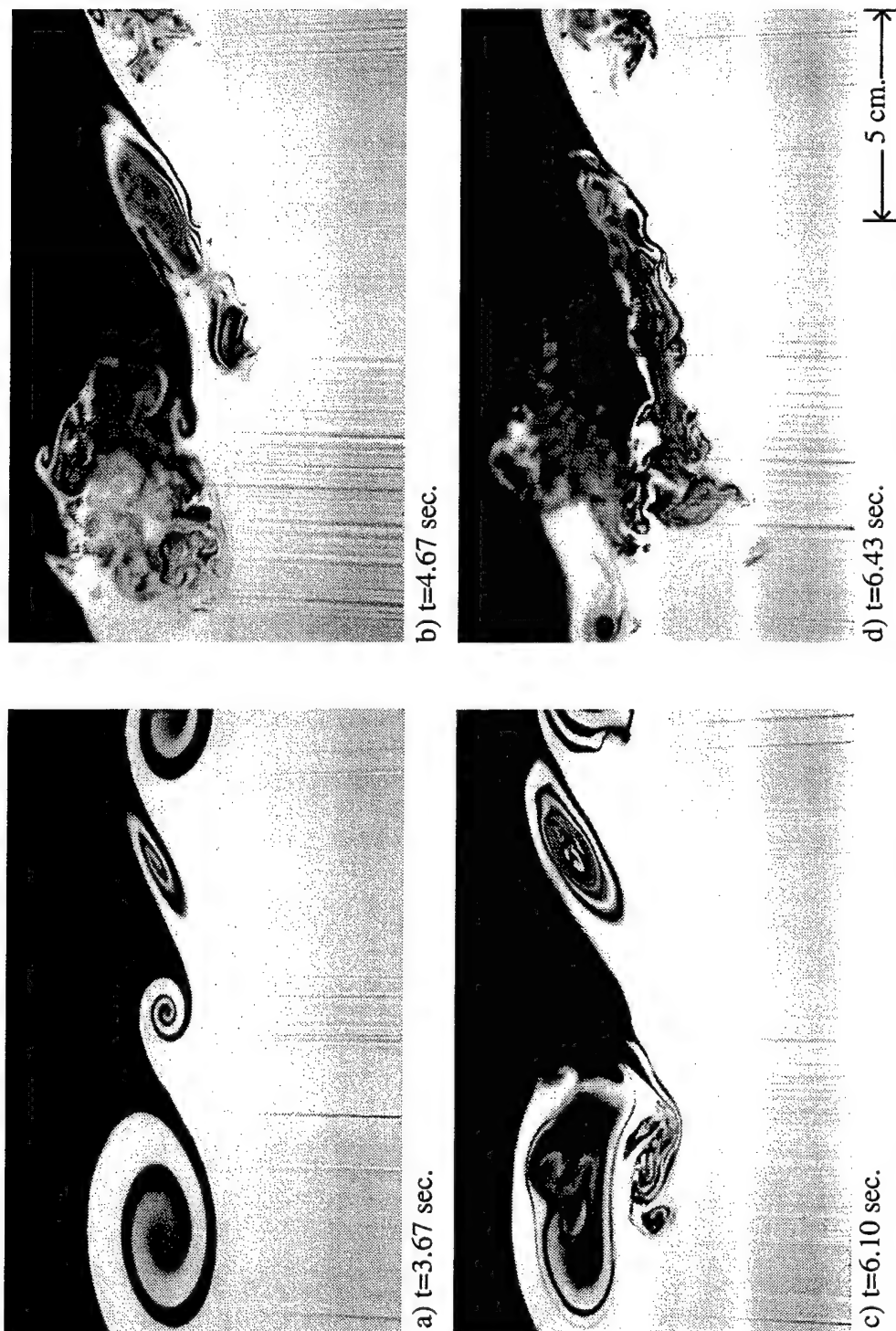


FIGURE 14. Side view showing vortex tearing in a stratified mixing layer with $Re=1900$ and $Ri=0.048$

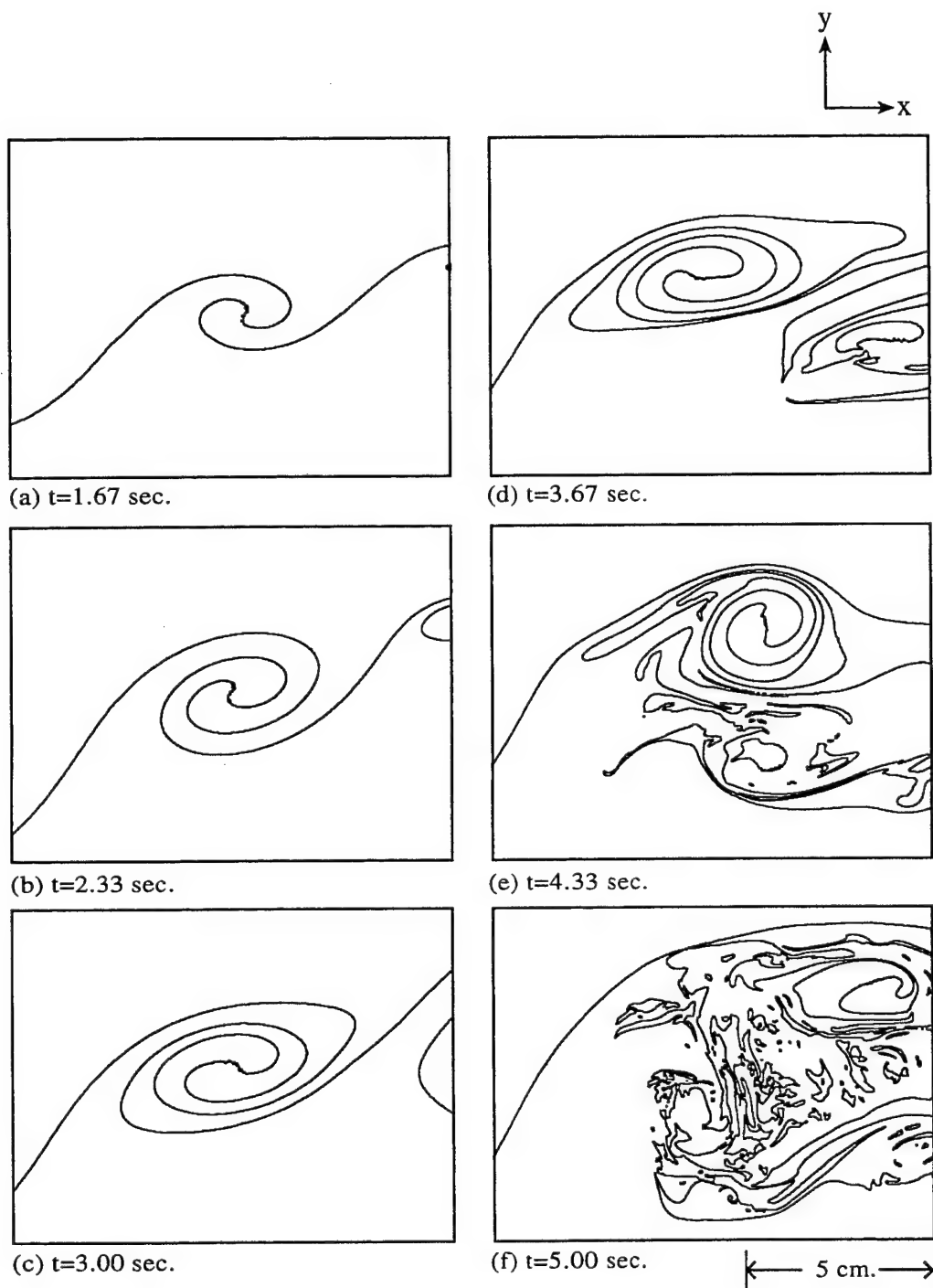


FIGURE 15. Images showing the material interfaces where diffusion is active. These images are equivalent to samples from figure 6 where $Re=2150$ and $Ri=0.012$. The time label for each image is from the onset of the Kelvin-Helmholtz instability. Scale is at the bottom right of the image.

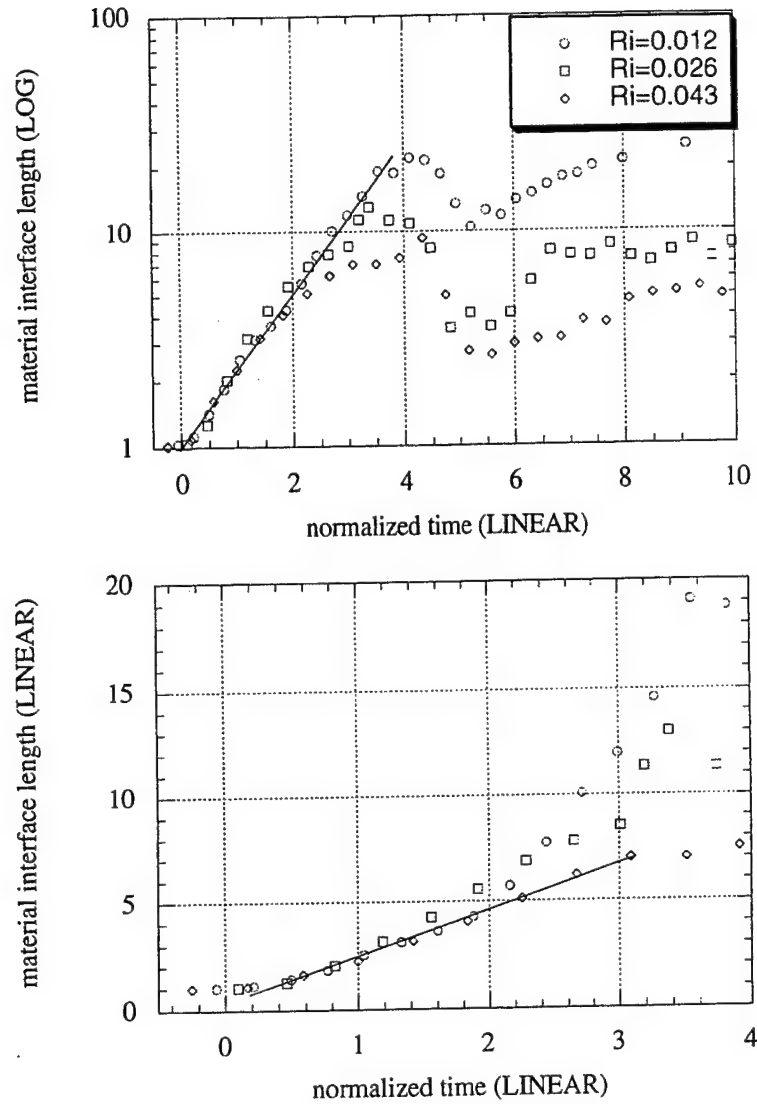


FIGURE 16. Material interface generation in the evolution of mixing layers. The time is normalized with the shear layer characteristic time $\lambda/(\Delta U/2)$.

in the appearance of an increasing number of folds of highly stretched material surfaces which elongate in one direction while contracting accordingly in the perpendicular direction. In the $Ri=0.012$ case the material interface stretches to more than 20 times its original length before the onset of small-scale three-dimensional motions (figure 6j). Since the recording system does not resolve the smallest scale in the flow, it is possible that the material-interface growth is significantly higher than measured. In figure 16 it can be seen that growth of the material interface for the two cases with vortical pairings does indeed follow an exponential line, whose slope of about 0.8 is the finite time Lyapunov exponent defined by

$$\beta = \frac{1}{\tau} \ln[L(\tau)/L(0)]$$

where L is the measured interfacial length and τ is the time in seconds normalized with the shear layer characteristic time $\lambda/(\Delta U/2)$. The material-interface growth for the higher Ri case of 0.043, however, appears to be linear (figure 16b). The main difference is that the vortices in this case do not go through pairings as in the two lower Ri cases.

After the onset of three-dimensional small-scale motions, a high level of mixedness is observed not only within the cores of the Kelvin-Helmholtz vortices but also in the area between two interacting large-scale structures where the complex stretches and folds were abundant. The present experimental observations serve to highlight some crucial steps in the chain of events that leads to an efficient mixing transition. The pairing kinematics of the two-dimensional large-scale structures could be important mixing processes, not only by creating a large increase in the diffusive interface through which fluids can mix in the molecular sense but also by creating a system of thin stretches and folds which would easily break up and create even more diffusive interface in the presence of small scale three-dimensional motions associated with turbulent transition.

The dips in the material-interface growth seen in the graphs of figure 16 are due to the fact that the onset of small-scale transition in the Kelvin-Helmholtz vortices at the peak of the curve tends to smear out the sharp interfaces created in the rollup and chaotic stretching processes. At this point in time, the level of mixedness in the mixing layer is observed to increase sharply.

3.4 *Mixing measurements*

To determine the overall effect of stratification on the growth of the mixing layer, the 'total' amount of mixed fluid is measured for different Richardson numbers keeping the Reynolds number roughly constant. The procedure, as described in sections 2.5 and 2.6, consists of taking mixing measurements from LIF images after the mixing layer has gone through the turbulent transition, and the flow has relaminarized. The obtained images are corrected for the horizontal Gaussian profile dropoff and vertical laser dye attenuation before the amount of mixed fluid is calculated, as described in Appendix 5.1. The measured value would then represent the total amount of mixed fluid attained in a stratified mixing layer after the self-similar state of asymptotically zero growth rate is established. Any subsequent mixing due to diffusion through the restratified layer is comparatively insignificant.

Figure 17 shows that the ability of the mixing layer to mix is drastically reduced as the Richardson number is increased. The reason for this reduced mixing is a combination of the reduction in the fluid entrainment into the Kelvin-Helmholtz vortices (figure 10), the smaller entrainment ratio of the heavier bottom fluid into the cores (figure 9), the reduced frequency of pairing of the large-scale structures (discussed in section 3.1), and the subsequent arrest of turbulence as the stabilization effect of the stratification eventually relaminarizes the flow. Plotted along with the data of the present investigation are the comparable data from Koop & Browand (1979), who performed their experiments in a water tunnel and measured the fluid mixing, with a conductivity probe, at their last measuring location, presumably after the mixing layer has reached an self-similar state and does not produce significantly

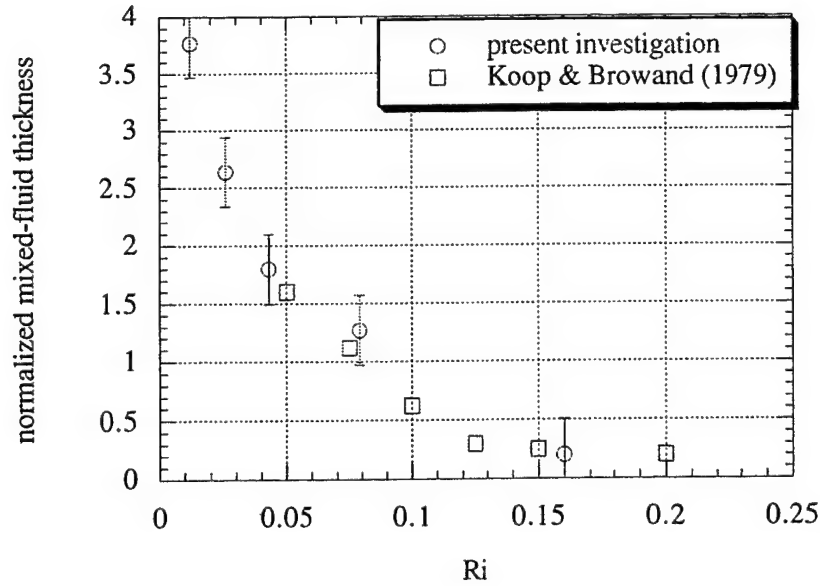


FIGURE 17. The mixed-fluid thickness normalized by the velocity thickness δ_v , vs. Richardson number. The Reynolds numbers for the experiments are all around 2000.

more mixing.

It is hard to gauge exactly what share of the decreased mixing each mechanism is responsible for as stratification is increased. It might be safe to say, however, that between $Ri=0.012$ and $Ri=0.043$, where the drop in mixing is sharpest, the main effect is the reduction in pairing and the corresponding reduction in material-interface stretching. The size of the Kelvin-Helmholtz vortices remains roughly constant within that Richardson number range (figure 10), and the entrainment ratio only decreases from about 1 to 0.8; these two effects would only account for a small reduction in mixing while the mixing is actually reduced by more than two folds. Perhaps not coincidentally, the peak interfacial length also decreases a little more than two folds between $Ri=0.012$ and $Ri=0.043$ (figure 16). Between 0.043 and 0.16, however, there is very little pairing but a four-fold reduction in the size of the large-scale vortices and a two-fold reduction in their entrainment ratio. Therefore, within this range, the reduction in mixing, about seven folds, seems to be due to the stabilizing effect on the entrainment of the large-scale vortices as stratification is increased.

4. Conclusions

Shear and buoyancy are the two main forces that determine structures and flow dynamics in a turbulent stratified mixing layer. The shear-induced features of the flow, though dominant, are affected in major ways by the presence of even a moderate amount of buoyancy. Stable stratification, commonly known for having a stabilizing effect on shear flows, reduces the size of the main Kelvin-Helmholtz vortices, resulting not only in smaller entrainment but also an entrainment ratio that is less favorable to efficient mixing. More importantly, stratification reduces pairing interactions among the large-scale vortices. Large-scale vortex pairing has for some

time been identified with the mixing transition. Winant & Browand (1974) observed through flow visualizations that small-scale disturbances appear some time after the first pairing. Jimenez (1983) associated the mixing transition with the emergence of the inertial subrange in the power spectra of velocity fluctuations, which coincides with the fourth pairing in his experiments.

The results from this investigation shed some light on the association between the mixing transition and large-scale vortex pairings. The main prerequisite to a mixing transition is a large increase in the diffusive interface where the two fluids are brought into direct contact so they can mix in the molecular sense; this is particular true in aqueous flows where the molecular diffusivity is very small, and diffusive mixing is a slow process. Meeting this requirement is generally attributed to the onset of small-scale three-dimensional motions through a cascade of secondary two-dimensional and three-dimensional instabilities such as streamwise counter-rotating vortices (Bernal & Roshko, 1986) and convective instability (Thorpe, 1971) in the stratified case. But even before the onset of small-scale transition, the diffusive interface between the top and bottom fluids can exponentially increase to more than 20 folds by complex kinematics during the first pairing of the Kelvin-Helmholtz vortices. This behavior not only contributes to the required diffusive interface directly, but the complex stretches and folds of the material interface effectively 'stir' the fluids so that once the small-scale transition takes place the amount of diffusive interface attained is greatly multiplied, resulting in a more efficient mixing transition. There has been a few previous works which generally support this observation. Jimenez & Martel (1991) observed a transition during the first pairing in a numerical two-dimensional mixing layer where the fractal geometry of the material interface increases from a low initial dimension of about 1-1.2 to a higher one of about 5/3. Min (1994) observed exponential line stretching of the interface in a two-dimensional mixing layer after the second pairing, indicating a chaotic behavior.

By reducing the size and entrainment of the large-scale Kelvin-Helmholtz vortices and deterring their pairing interactions, the presence of stable stratification has a net effect of reducing the overall mixing. But there are two mechanisms associated with buoyancy which actually has quite the reverse effect. Through baroclinic vorticity generation in the braid and the strain field of the main Kelvin-Helmholtz vortices, a secondary Kelvin-Helmholtz-like instability usually develops given an adequately-low local Richardson number. In fact this type of instability is not limited to the braid but also sometimes occurs on top of a large-scale vortex provided a high enough local shear. Secondly, the statically-unstable regions inside the cores of the large-scale vortices can support gravitational instabilities in the forms of two-dimensional rolls or three-dimensional fingers. These two secondary instabilities would naturally add to the overall mixing in the mixing layer, but the effect is only secondary compared with the stabilizing effect of stratification.

Other secondary instabilities such as vortex dislocations and tubes & knots occur sporadically throughout the mixing layer. They contribute to the small-scale processes because they have a good streamwise vorticity component which is subjected to stretching by the strain field between two large-scale vortices and the general strain field of the mixing layer. Vortex stretching is the primary turbulence maintaining mechanism in a free shear flow. Structures that are subject to a high strain rate (such as in the braid, along the principle axis of strain), are especially efficient in contributing to the small-scale generation.

We sincerely thank Professors A. Roshko, F. K. Browand, C. D. Winant, P. E. Dimotakis, and A. Leonard for the valuable discussions which contributed a great deal to this work. This work was supported by DARPA grant N00014-91-J-1968.

5. Appendices

5.1 Statistical measurements from LIF measurements

From the density measurements, we can obtain statistical quantities about the mixed turbulent fluids such as the p.d.f. of the bottom-layer-fluid concentration as a function of y . From the p.d.f. we can calculate various quantities such as the average concentration of the bottom-layer fluid as a function of y , the probability of finding mixed fluid of any concentration as a function of y , and the average mixed fluid concentration as a function of y (Koochesfahani & Dimotakis, 1986).

We denote the bottom-layer-fluid concentration ξ where ξ is defined as the bottom-layer fluid volume fraction $v_b/(v_t+v_b)$. Normalization of the p.d.f. $P(\xi, y)$ requires that

$$\int_{-1}^1 P(\xi, y) d\xi = 1$$

Concentrations in the range $0 \leq \xi \leq \varepsilon$ are assigned to pure (unmixed) top-layer fluid and $1-\varepsilon \leq \xi \leq 1$ to pure bottom-layer fluid.

The average concentration of top-layer fluid $\bar{\xi}(y)$ is given by

$$\bar{\xi}(y) = \int_0^1 \xi P(\xi, y) d\xi$$

The probability of finding mixed fluid at any concentration $P_m(y)$ is given by

$$P_m(y) = \int_{\varepsilon}^{1-\varepsilon} P(\xi, y) d\xi$$

The average mixed-fluid concentration $\bar{\xi}_m(y)$ is given by

$$\bar{\xi}_m(y) = \frac{\int_{\varepsilon}^{1-\varepsilon} \xi P(\xi, y) d\xi}{\int_{\varepsilon}^{1-\varepsilon} P(\xi, y) d\xi} = \frac{\int_{\varepsilon}^{1-\varepsilon} \xi P(\xi, y) d\xi}{P_m(y)}$$

A useful quantity called mixedness, $M(t, y)$ (Konrad 1976) is defined by

$$M(t, y) = \frac{\int_{x_1}^{x_2} H(\rho - \bar{\rho}) (\rho_b - \rho) dx + \int_{x_1}^{x_2} H(\bar{\rho} - \rho) (\rho - \rho_t) dx}{\int_{x_1}^{x_2} H(\rho - \bar{\rho}) (\rho_b - \bar{\rho}) dx + \int_{x_1}^{x_2} H(\bar{\rho} - \rho) (\bar{\rho} - \rho_t) dx}$$

where x_1 and x_2 are the edges of the region of interest, and H is the Heaviside step-function defined as

$$H(f) = 1 \quad \text{for } f \geq 0$$

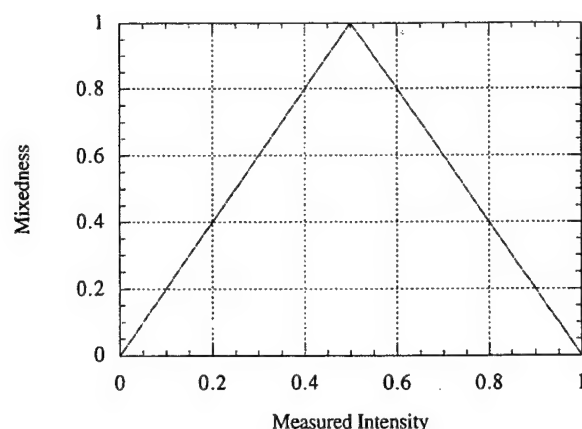


FIGURE A1: Mixedness calculated from measured fluorescent intensity of each measuring volume

$$= 0 \quad \text{for } f < 0$$

The mixedness is zero when the fluid in the measuring volume registers either 0 or 1 in the fluorescent intensity measurement, meaning that the measuring volume has no mixed fluid, containing only of either fluid from the top or from the bottom layer. The mixedness would be one if the measuring volume records an intensity of 0.5, meaning that there is an equal amount of top and bottom fluid in the measuring volume and that the measuring volume consists entirely of mixed fluid, provided the spatial resolution of the recording system can resolve the smallest diffusion scale. Figure A1 depicts the mixedness as a function of measured intensity.

The mixedness is also physically related to the amount of product that would be created if the top and the bottom layers each contains a small and equal concentration of reactants A and B, respectively; and the chemical reaction is one-step, fast, and irreversible with molar stoichiometric coefficients of 1:



In fact, the amount of product would be equal to half the mixedness:

$$\text{Product}(t,y) = 1/2 M(t,y)$$

We define a mixing layer density thickness L_p as the width between the two vertical points where $P_m(y)$ has fallen to 1% of its maximum value. This width is approximately equal to the visual width of the mixing layer.

5.2 Calculation of flow parameters

The flow parameters such as speed and viscous diffusion length scale can be analytically derived from fundamental equations. From these calculations, flow parameters such as the Reynolds number and the Richardson number can be estimated.

Consider the two-layered system initially at rest which at time $t=0$ is tilted to the angle θ . For the parallel accelerating flow that follows, the inviscid equations of motion that describes the flow before the initial instability are:

$$\rho \frac{\partial u}{\partial t} = -\frac{\partial p}{\partial x} - g\rho \sin \theta$$

and

$$0 = -\frac{\partial p}{\partial y} - g\rho \cos \theta$$

Applying continuity, it can be shown that

$$u(y,t) = gt \sin \theta \left(\frac{H}{\rho(y) \int_{-H/2}^{H/2} \frac{dy}{\rho}} - 1 \right), \quad (\text{Thorpe, 1968})$$

where H is the height of the tank.

The free stream velocities of the two layers, therefore, can be written as:

$$U_t = \frac{\Delta \rho g t \sin \theta}{2\bar{\rho}}, \quad \text{and} \quad U_b = -\frac{\Delta \rho g t \sin \theta}{2\bar{\rho}}.$$

and the velocity difference between the two layers as:

$$\Delta U = U_t - U_b = \frac{\Delta \rho}{\bar{\rho}} g t \sin \theta$$

Thorpe (1985) has calculated the ratio of the velocity gradient at $y = 0$ in a viscous flow to that at $y = 0$ in an inviscid flow, Q , for the flow resulting from the density profile

$$\rho(y) = \rho_{av} \left[1 - \frac{\Delta \rho}{2} \operatorname{erf}(y\sqrt{\pi}/\delta_\rho) \right]$$

The ratio Q is equal to the ratio δ_ρ/δ_v and has the form

$$Q = \frac{2}{1 + \left(1 + \frac{4\pi\nu t}{\delta_\rho^2} \right)^{1/2}}$$

The flow Reynolds number and Richardson number are, therefore, approximately given by:

$$\operatorname{Re} = \frac{\Delta U \delta_\rho}{\nu Q} = \left(\frac{\Delta \rho}{\bar{\rho}} g t \sin \theta \right) \frac{\delta_\rho}{\nu Q}$$

$$\operatorname{Ri} = \frac{g \Delta \rho \delta_\rho}{\bar{\rho} (\Delta U)^2 Q^2}$$

REFERENCES

- Aref, H. 1984 Stirring by chaotic advection. *J. Fluid Mech.* **143**, 1.
- Bernal, L. P. 1988 The statistics of the organized vortical structure in turbulent mixing layers. *Phy. Fluids A*, **31**, 2533.
- Bernal, L. P. & Roshko, A. 1986 Streamwise vortex structure in plane mixing layers. *J. Fluid Mech.* **170**, 499.
- Breidenthal, R. 1978 Chemically reacting, turbulent shear layer. Ph.D. thesis, California Institute of Technology.
- Breidenthal, R. 1981 Structure in turbulent mixing layers and wakes using a chemical reaction. *J. Fluid Mech.* **109**, 1.
- Browand, F. K. & Ho, C.-M. 1987 Forced, unbounded shear flows. *Nuclear Physics B*, **2**, 139.
- Browand, F. K. & Troutt, T. R. 1980 A note on spanwise structure in the two-dimensional mixing layer. *J. Fluid Mech.* **97**, 771.
- Brown, G. L. & Roshko, A. 1974 On density effects and large structure in turbulent mixing layers. *J. Fluid Mech.* **64**, 775.
- Browning, K. A. & Watkins, C. D. 1970 Observations of clear air turbulence by high power radar. *Nature* **227**, 260.
- Chandrsuda, C., Mehta, R. D., Weir, A. D. & Bradshaw, P. 1978 Effect of free stream turbulence on large structure in turbulent mixing layers. *J. Fluid Mech.* **85**, 693.
- Chien, W.-L., Rising, H. & Ottino, J. M. 1986 Laminar mixing and chaotic mixing in several cavity flows. *J. Fluid Mech.* **170**, 355.
- Corcos, G. M. & Sherman, F. S. 1984 The mixing layer: deterministic models of a turbulent flow. Part 1. Introduction and the two-dimensional flow. *J. Fluid Mech.* **139**, 29.
- Dimotakis, P. E. & Brown, G. L., 1976 The mixing layer at high Reynolds number: large-structure dynamics and entrainment. *J. Fluid Mech.* **78**, 535.
- Davey, R. F. & Roshko, A. The effect of a density difference on shear-layer instability. *J. Fluid Mech.* **53**, 523.
- Fritts, D. C. & Rastogi, P. K. 1985 Convective and dynamical instabilities due to gravity wave motions in the lower and middle atmosphere: theory and observations. *Radio Science* **20**, 1247.
- Hazel, P. 1969 Ph.D. thesis, University of Cambridge.
- Hazel, P. 1972 Numerical studies of the stability of inviscid stratified shear flows. *J. Fluid Mech.* **51**, 39.
- Ho, C.-M. & Huerre, P. 1984 Perturbed free shear layers. *Ann. Rev. Fluid Mech.* **16**, 365.
- Ho, C.-M. & Nosseir, N. 1981 Dynamics of an impinging jet. Part 1 The feedback phenomenon. *J. Fluid Mech.* **105**, 119.
- Hussain, A. K. M. F. 1986 Coherent structures and turbulence. *J. Fluid Mech.* **173**, 303.
- Jimenez, J. 1983 A spanwise structure in the plane shear layer. *J. Fluid Mech.* **132**, 319.
- Jimenez, J. & Martel C. 1991 Fractal interfaces and product generation in the two-dimensional mixing layer. *Phy. Fluids A*, **3**, 5.
- Klaassen, G. P. & Peltier, W. R. 1985 The onset of turbulence in finite-amplitude Kelvin-Helmholtz billows. *J. Fluid Mech.* **155**, 1.
- Konrad, J. H. 1977 An experimental investigation of mixing in two-dimensional turbulent shear flows with applications to diffusion-limited chemical reactions. Ph.D. thesis, California Institute of Technology.
- Koochesfahani, M. M. & Dimotakis, P. E. 1986 Mixing and chemical reactions in a turbulent liquid mixing layer. *J. Fluid Mech.* **170**, 83.
- Koop, C. G. & Browand, F. K. 1979 Instability and turbulence in a stratified fluid with shear. *J. Fluid Mech.* **93**, 135.
- Lasheras, J. C. & Choi, H. 1988 Three-dimensional instability of a plane free shear layer: an experimental study of the formation and evolution of streamwise vortices. *J. Fluid Mech.* **189**, 53.

- Liepmann, H. W. & Laufer, J. 1947 Investigation of free turbulent mixing. *N. A. C. A. Tech. Note*, no. 1257.
- Lin, S. J. & Corcos, G. M. 1984 The mixing layer: deterministic models of a turbulent flow. Part 3. The effect of plane strain on the dynamics of the streamwise vortices. *J. Fluid Mech.* **141**, 139.
- Martin, J. E. & Meiburg, E. 1991 The three-dimensional evolution of axisymmetric jets perturbed by helical waves. *8th Symposium on Turbulent Shear Flows*, Technical University of Munich, 6-3-1.
- McDougall, T. J. 1979 On the elimination of refractive-index variations in turbulent density-stratified liquid flows. *J. Fluid Mech.* **93**, 83.
- Miksad, R. W. 1972 Experiments on the nonlinear stages of free-shear layer transition. *J. Fluid Mech.* **56**, 645.
- Miles, J. W. & Howard, L. N. 1964 Note on heterogeneous shear flow. *J. Fluid Mech.* **20**, 331.
- Min, I. A. 1994 Transport, stirring and mixing in two-dimensional vortex flows. Ph.D. thesis, California Institute of Technology.
- Moore, D. W. & Saffman, P. G. 1975 The density of organized vortices in a turbulent mixing layer. *J. Fluid Mech.* **69**, 465.
- Patnaik, P. C., Sherman, F. S. & Corcos, G. M. 1976 A numerical simulation of Kelvin-Helmholtz waves of finite amplitude. *J. Fluid Mech.* **73**, 215.
- Pierrehumbert, R. T. & Widnall, S. E. 1982 The two- and three-dimensional instabilities of a spatially periodic shear layer. *J. Fluid Mech.* **114**, 59.
- Roshko, A. 1991 The mixing transition in free shear flows. *The Global Geometry of Turbulence*, ed. J. Jimenez, Plenum Press, New York, 3.
- Schowalter D. G. 1993 The effect of stable stratification on three-dimensional structure in shear layers. Ph.D. thesis, University of California at San Diego.
- Staquet, C. & Riley J. J. 1989 A numerical study of a stably-stratified mixing layer. *Turbulent Shear Flows 6*. Springer-Verlag, 381-397.
- Thorpe, S. A. 1968 A method of producing a shear flow in a stratified fluid. *J. Fluid Mech.* **32**, 693.
- Thorpe, S. A. 1971 Experiments on the instability of stratified shear flows: miscible fluids. *J. Fluid Mech.* **46**, 299.
- Thorpe, S. A. 1973 Experiments on instability and turbulence in a stratified shear flow. *J. Fluid Mech.* **61**, 731.
- Thorpe, S. A. 1985 Laboratory observations of secondary structures in Kelvin-Helmholtz billows and consequences for ocean mixing. *Geophys. astrophys. Fluid Dynamics* **34**, 175.
- Thorpe, S. A. 1987 Transitional phenomena and the development of turbulence in stratified fluids: a review. *J. Geop. Res.* **92**, 5231.
- Winant, C. D. & Browand, F. K. 1974 Vortex pairing: the mechanism of turbulent mixing-layer growth at moderate Reynolds number. *J. Fluid Mech.* **63**, 237.
- Woods, J. D. 1968 Wave-induced shear instability in the summer thermocline. *J. Fluid Mech.* **32**, 791.

Structures in stratified mixing layers Part 2. The effects of cross shear

By P. ATSAVAPRANEE AND M. GHARIB

Graduate Aeronautical Laboratories, California Institute of Technology,
Pasadena, CA 91125, USA

In the second part of this study, the effect of cross shear on a nominally two-dimensional plane mixing layer is investigated. A temporal mixing layer is generated with a stratified tilting tank similar to that used by Thorpe [J. Fluid Mech. 32, 693 (1968)]; and cross shear, with the vorticity component in the streamwise direction, is introduced to this mixing layer to simulate situations that are widespread in industrial and natural flows. A new type of "co-rotating" streamwise vortices arises in the braid region of the mixing layer from the introduction of cross shear. The appearance and dynamics of these secondary co-rotating streamwise vortices are observed to be very similar to those of the primary Kelvin-Helmholtz instability, exhibiting similar roll-up and pairing, both qualitatively and quantitatively in the normalized length scale (wavelength) and time scale (time of pairing).

Measurements of the density field by the laser-induced fluorescence technique (LIF) show that mixing increases appreciably when a critical level of cross shear is introduced to the mixing layer, inducing the co-rotating streamwise vortices. The facts that the streamwise vortices reside in the braid and are roughly aligned along the principle axis of strain subject them to large vortex stretching, which means they are able to generate a considerable amount of turbulent stresses and small-scaled fluctuations. It is also found that if cross shear is introduced while the mixing layer is still essentially a plane two-dimensional shear layer, the mixing layer simply behaves as a normal two-dimensional mixing layer, with the flow redirected along an oblique axis. The co-rotating streamwise vortices and the corresponding increased mixing are produced only when cross shear is introduced around the time the primary shear layer starts to roll up into Kelvin-Helmholtz vortices and is no longer a plane two-dimensional shear layer. Evidence also shows that even in a two-dimensional stratified mixing layer without "global" cross shear, the co-rotating streamwise vortices can develop where local cross shear is induced by streamwise baroclinic vorticity generation at locations with a high local curvature of the density interface and a large $\partial\rho/\partial z$ component.

1. Introduction

Most of the previous studies in turbulent mixing layers have been done on the relatively simple geometry of the homogeneous two-dimensional shear flow for several reasons, the obvious being that this geometry represents the most generic shear flow, the knowledge of which studies of more complex flows can draw upon. Gradually, research in turbulent mixing layers is maturing into more complex flows in order to model more complicated and realistic situations, building upon the foundation of previous efforts. Natural and industrial mixing layers involve complexities of many different kinds. One such complexity that deserves attention

but has so far not been studied more thoroughly, is the non-steady nature of the base flow, the freedom of the base flow to vary in magnitude and direction with time. The flow dynamics of a mixing layer that involves this kind of complexity would be inherently different from that of a mixing layer whose base flow is constrained to be two-dimensional, uni-directional, and steady. A likely scenario in the ocean or the atmosphere, for example, would be that while a two-dimensional mixing layer is developing, the base flow could easily shift into an oblique direction (in the original plane), effectively adding another component of shear to the mixing layer.

With the introduction of cross shear to the two-dimensional mixing layer, the obvious question that arises is not if but in what way the introduction of cross shear would change the overall dynamics of the mixing layer. Would the general behavior of the spanwise Kelvin-Helmholtz instability be affected? Would the counter-rotating streamwise vortices of the kind studied by Bernal & Roshko (1986) be enhanced because streamwise vorticity is being introduced or would another type of streamwise instability result instead? What are the effects of cross shear on the turbulent transition and the subsequent turbulent state?

2. Experimental apparatus and methods

The experimental apparatus has been described in detail in the first part of this two-part series (Atsavapranee and Gharib, 1995), heretofore referred to as Part I. An enclosed tank, 96" long, 30" wide, and 5" deep, is filled with two layers of stratified fluids: in most cases, epsom salt solution on the bottom and ethanol solution on top, with the index of refraction of the two layers matched for the purpose of optical measurements. Tilting the tank generates the potential for the two layers to accelerate in opposite directions, creating shear at the interface. The tank is designed so that it can be tilted around two independent axes of rotations. The tank is first tilted around the main axis, generating the main shear; and a moment later, with the primary shear layer still evolving, the tank is tilted around the secondary axis, for the cross shear. The resulting shear layer is not simply a two-dimensional layer with slanted vortices if the cross shear is introduced after the rollup process of the Kelvin-Helmholtz instability starts. The flow is uniquely determined by these non-dimensional parameters:

$$Ri_i = g\Delta\rho\delta/\rho_{av}(\Delta U)^2; \quad Re_i = \Delta U\delta/\nu$$

$$Sc = \nu/D; \quad \delta/\delta_p; \quad \Delta U_c/\Delta U_m; \quad \Delta T/(\delta/\Delta U)$$

where ν and D are the kinematic viscosity and the molecular diffusivity of water, respectively. Ri_i and Re_i are the initial Richardson number and Reynolds number, measured at the onset of the Kelvin-Helmholtz instability, respectively. Sc is the Schmidt number which is about 600. ΔU_c and ΔU_m are the cross shear and main shear velocity differences between the top and bottom streams. And ΔT is the time lag between the main and the cross shear.

Two-dimensional planes in the flow are visualized using laser-induced fluorescent dye. Sodium fluorescein (uranine) is premixed with the bottom layer fluid and its fluorescence is excited by an argon-ion laser sheet. The plan view of the mixing layer is visualized in two ways. The first method is the standard shadowgraph technique, and the second method is similar to that used by Lasheras and Choi (1988). The interface between the two streams can be visualized by premixing the bottom layer with a high concentration of a fluorescent dye and illuminating from above with flash lamps to create an opaque interface. The two-dimensional plane

LIF images, after being corrected for laser sheet non-uniformity and laser dye attenuation, are used to calculate the two-dimensional density field of the mixing layer, in a way similar to the line density measurements by Koochesfahani & Dimotakis (1986). "Overall" mixing attained by the mixing layer is measured by the LIF density field measurement technique.

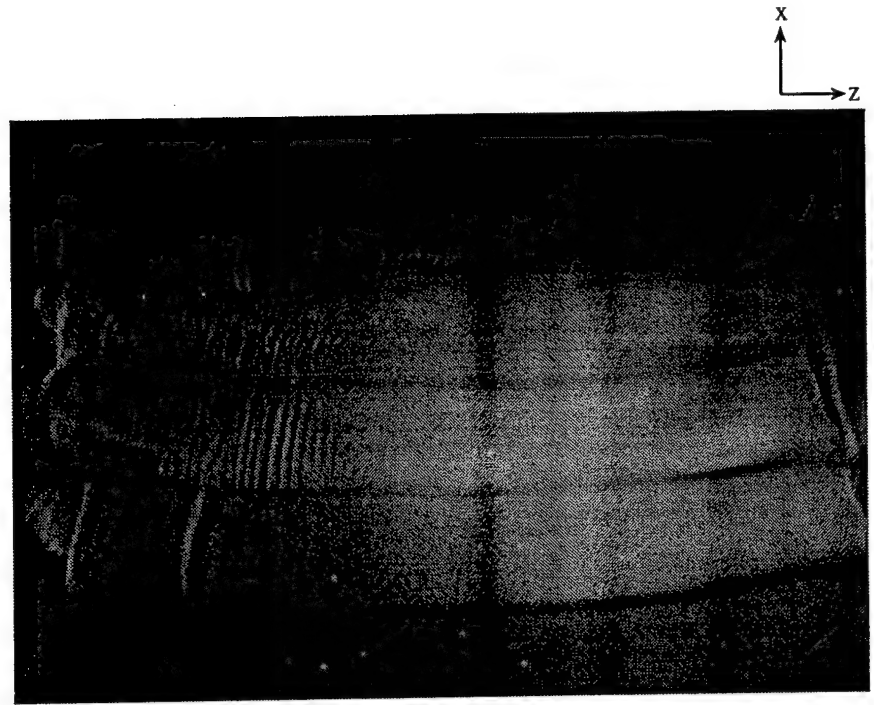
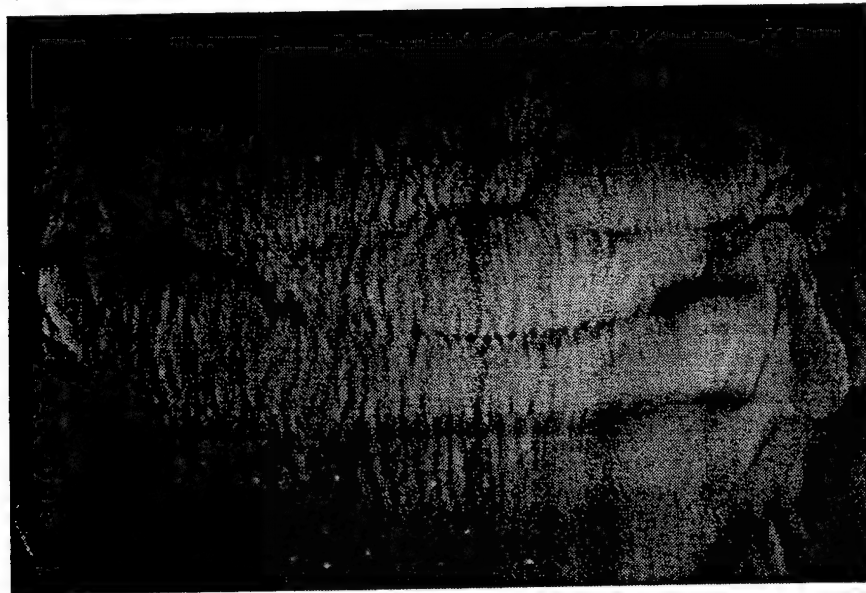
3. Results and discussions

3.1 Flow Visualizations

Figure 1 illustrates the plan view of a mixing layer with cross shear. In these photographs essentially a thin layer on the interface of the top and bottom layers is visualized by premixing the bottom layer with a high concentration of a fluorescent dye and illuminating from above with flash lamps to create an opaque interface. The dark lines represent the edges of vortical structures that form shadows when the interface is illuminated from top. The main shear gives a vorticity component in the negative z direction, and the cross shear in the negative x direction. As expected, the primary Kelvin-Helmholtz instability amplifies as a result of the main shear, with the primary vortices aligned along the z -axis (spanwise direction). The introduction of the cross shear does not seem to alter the characteristics of the Kelvin-Helmholtz vortices in any significant way. But along the braid, very organized streamwise streaks which are roughly perpendicular to the main vortices are observed to develop (figure 1a). The length scale of these streaks, after they first appear, is initially about 0.75 cm., an order of magnitude lower than that of the primary vortices which have a wavelength of about 7.5 cm. As the mixing layer develops, the length scale of the streamwise streaks increases, seemingly going through successive doublings in scale (figure 1b). Eventually, the mixing layer goes through turbulent transition in which a large increase in random small-scale three-dimensional motions is observed.

Figure 2 is a plan-view shadowgraph of a stratified mixing layer to which cross shear is introduced just about when the first Kelvin-Helmholtz instability becomes visible. Again streamwise streaks are seen developing in the braid region between two Kelvin-Helmholtz vortices. Note also the helical pairing of the primary Kelvin-Helmholtz vortices, a three-dimensional instability mode of the spanwise vortices identified by Pierrehumbert and Widnall (1982). The y - z view or a cross-sectional view of the streamwise streaks illuminated by a laser sheet (figure 3) clearly shows that these streaks are in fact structures of concentrated vorticity and that they are of the same sign everywhere.

Qualitatively, the streamwise vortices appear and behave very much as the primary Kelvin-Helmholtz vortices in that they roll up and pair. Quantitatively, the length scale and time scale of the streamwise vortices are also roughly in the order of those of the primary vortices. We use as length scale the wavelength of the structures normalized by the local density thickness and as time scale the time it takes for two structures to pair normalized by the overturning time scale $\lambda/(\Delta U/2)$. The average ratio of the streamwise to primary length scales is 0.7, and the average ratio of the streamwise to primary time scales is 1.2. Figure 4 is a sequence of pictures showing the pairing of two streamwise vortices. It is normal for the streamwise vortices to pair twice or three times, eventually making their sizes comparable to those of the primary vortices which pair much slower. Figure 5 reveals a braid region being layered in alternate dark and bright bands as if in a spiral, indicating that there is a streamwise vortical structure present. The photo-

a) $t=1.2$ sec.b) $t=1.6$ sec.

← 15 cm. →

FIGURE 1: Plan view of a mixing layer to which a moderate amount of cross shear is imposed. Streamwise streaks develop along the braid and increase in scale as they evolve. (a) and (b) are 1.2 and 1.6 sec. after the onset of the primary Kelvin-Helmholtz instability. Cross shear is introduced 4 sec. after the main shear, shortly after the onset of the primary K-H instability (at 3.9 sec.). The Reynolds number of the main shear is 2800. The Richardson number is 0.034 for the main shear and 0.04 for the cross shear. The cross shear is about 21.5% of the main shear.

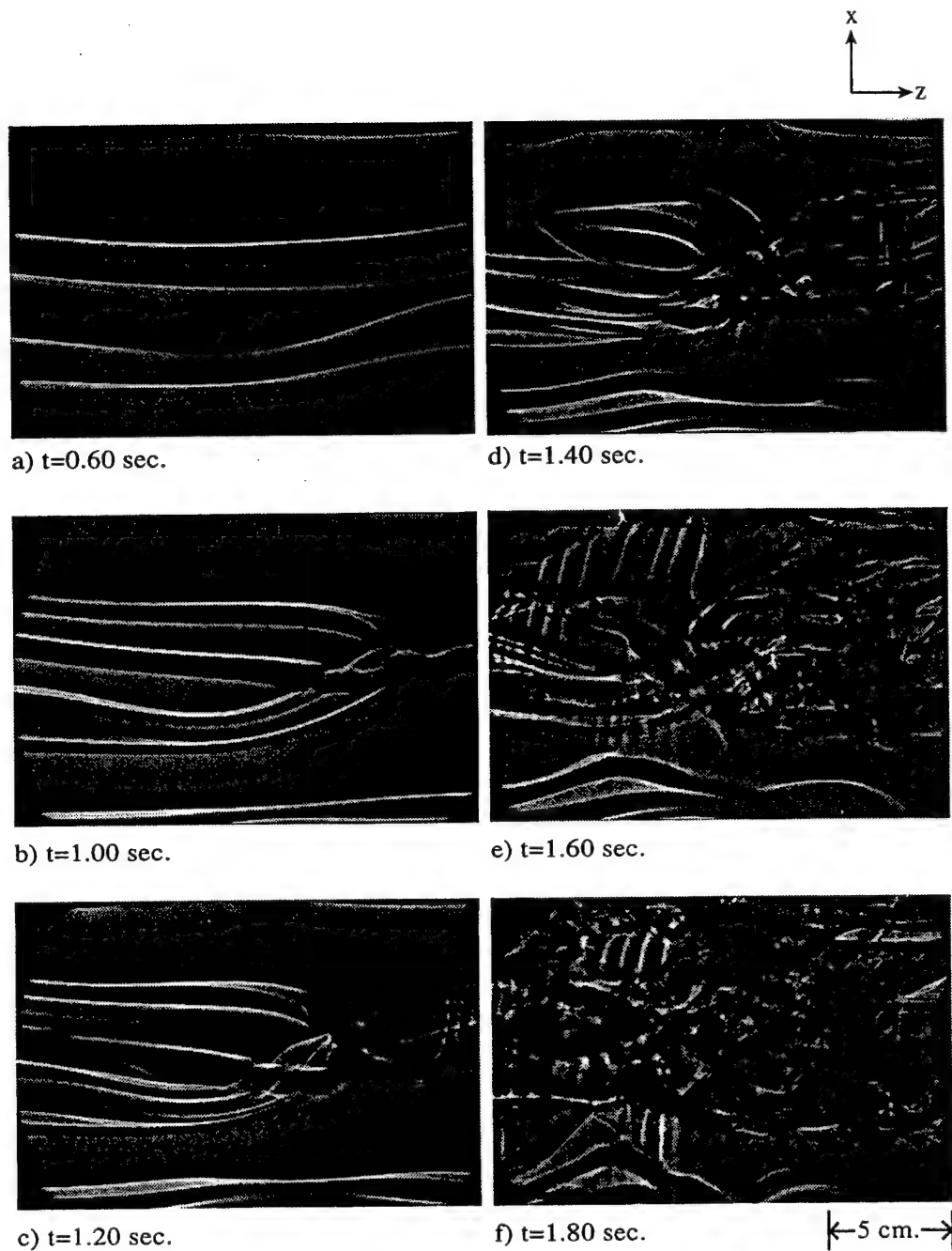


FIGURE 2: Plan-view shadowgraph of a stratified mixing layer with cross shear. $Re=1900$ and $Ri=0.048$. Cross shear is about 32.5% of the main shear. The time for each image is from the onset of the primary Kelvin-Helmholtz instability.



FIGURE 3: A cross-sectional view of the streamwise streaks showing the "co-rotating" streamwise vortices. The image is taken 1.50 seconds after the onset of the primary Kelvin-Helmholtz instability. Cross shear is introduced 4 secs after the main shear, about the same time as the onset of the primary Kelvin-Helmholtz instability (at 4.1 sec.). The Reynolds number of the main shear is 1880. The Richardson number is 0.09 for the main shear and 0.02 for the cross shear. The cross shear is approximately 50% of the main shear.

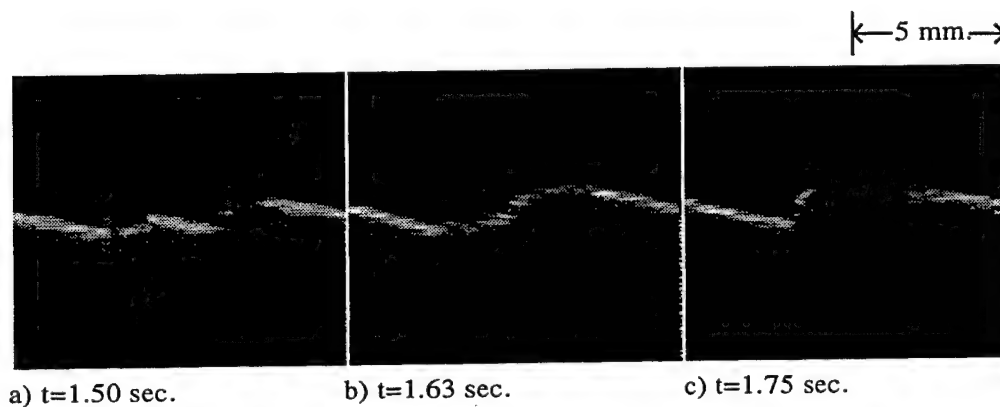


FIGURE 4: A sequence of pictures showing two "co-rotating" streamwise vortices in the process of pairing. The experiment is the very same one as in figure 3. The times of the sequence are after the onset of the primary K-H instability.

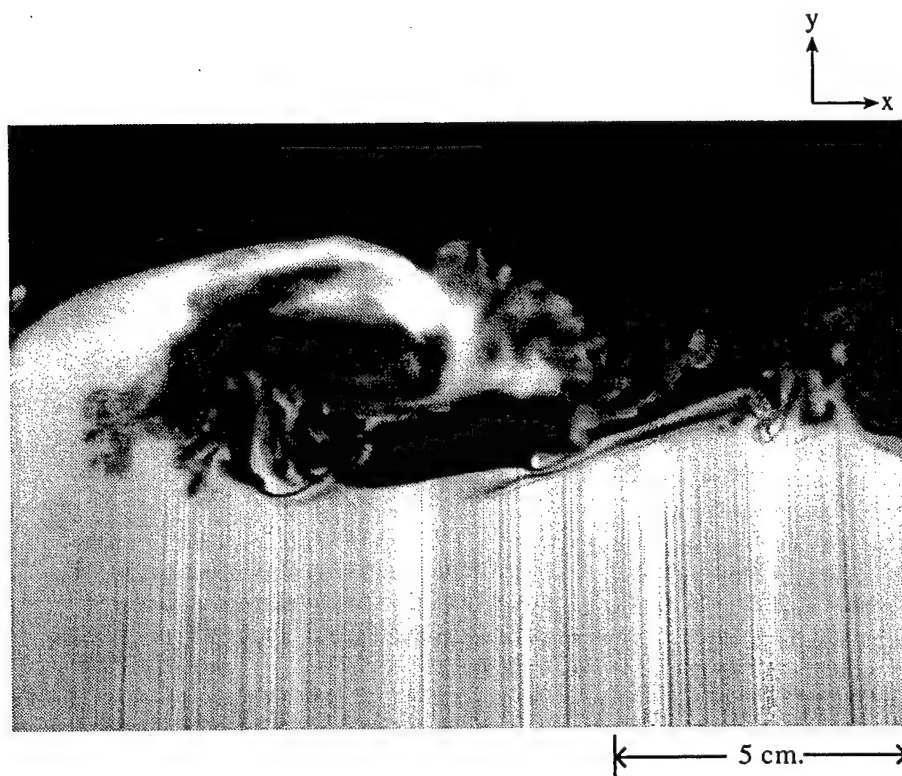


FIGURE 5: Side view of streamwise rollers in a mixing layer with $Re=1950$ and $Ri=0.046$. The cross shear is about 25% of the main shear.

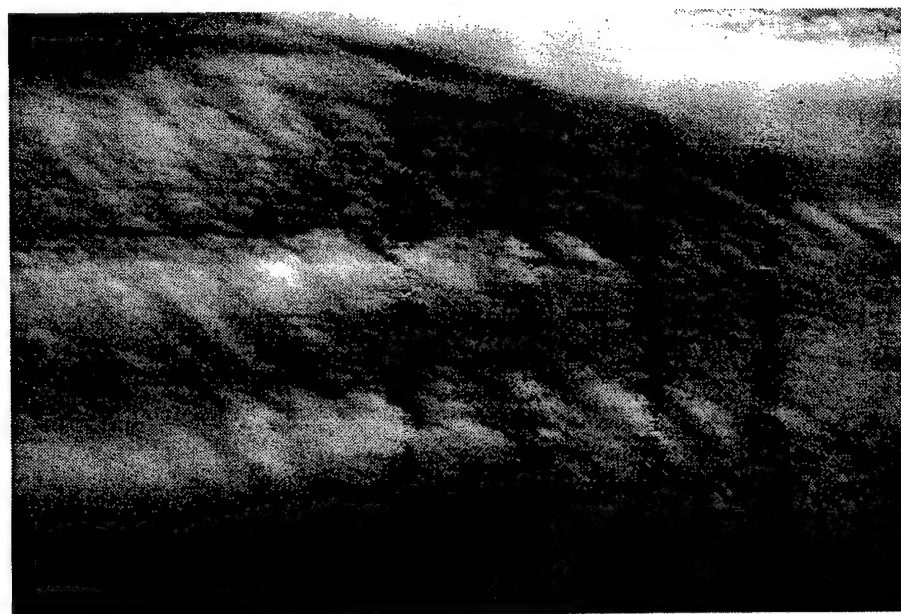


FIGURE 6: Cloud pattern showing large-scale spanwise rolls and small-scale longitudinal rolls similar to "co-rotating" streamwise vortices

graph of clouds in figure 6 displays a pattern with long wavelength bands in one direction, and smaller wavelength bands in the perpendicular direction. Without any scientific claim we raise a question here if it is possible for the same type of instability to exist in such a wide range of conditions.

3.2 *A simple heuristic flow model*

From these observations, it is clear that the streamwise streaks are in fact concentrated streamwise vortical structures and that they are direct consequences of the cross shear. A simple heuristic model of the dynamics can be argued from the available information. The introduction of cross shear to a developing plane mixing layer essentially has the effect of imposing streamwise vorticity everywhere in the mixing layer. Around the stagnation point in the braid region of the primary vortices, the local flow field looks like that of Hiemenz flow. This flow field creates a strain field that stretches and magnifies the streamwise vorticity layer in the braid which would very likely become unstable in a Kelvin-Helmholtz-like manner. The braid strain field would further stretch the resulting streamwise vortices, which may then go through pairings similar to normal Kelvin-Helmholtz vortices. The length scale of the streamwise vortices would naturally scale on the density thickness around the braid where they originate, which is about an order of magnitude lower than that of the primary shear layer.

The stability and numerical analyses of Lin and Corcos (1984) study the development of counter-rotating streamwise vortices in the strain field of the braid region between two large-scale Kelvin-Helmholtz vortices by modelling the local flow field around the stagnation point of the shear layer with a plane strain field and imposing on that system a spanwise-sinusoidal distribution of streamwise vorticity. In section 4 of that work, however, the authors study a special case in which the wavenumber of the initial sinusoidal streamwise vorticity goes to zero, meaning the imposed streamwise vorticity is basically a two-dimensional plane streamwise vortex sheet, a situation exactly the same as our conceptual model for the co-rotating streamwise vortices. They determine the stability characteristics of a strained vortex sheet and show that a strained vortex sheet exhibits rollup and pairing dynamics similar to those in the unstrained case. The conceptual model, treating the co-rotating streamwise vortices as strained Kelvin-Helmholtz vortices, simply ignores the question of how the ends of the streamwise vortices interact with the primary Kelvin-Helmholtz structures. Even then the measured streamwise to primary length-scale and time-scale ratios (section 3.1) are in line with the behavior of the instability in a strained vortex sheet which has a higher wavenumber and longer pairing time than the unstrained case (Lin and Corcos, 1984). The observed co-rotating streamwise vortices also have a close parallel in a helically-perturbed nominally-axisymmetric jet. Martin and Meiburg (1991) has observed Kelvin-Helmholtz-like streamwise vortices of the same sign in the braid regions of a numerical helical jet.

It is important to point out that the streamwise vortices appear even when the cross shear is very weak, implying that the roll-up of the streamwise vortex sheet into concentrated vortices are intimately related to the strain field imposed by the main Kelvin-Helmholtz structures. In fact, as in the experiment of figure 1, if the streamwise vortex sheet is present alone, it would not roll up since the Richardson number is about 0.4, well above the critical number of 0.25. But with the stretching action of the braid flow field, the local length scale of the braid can get drastically smaller, and the local Richardson number can be reduced by an order of magnitude

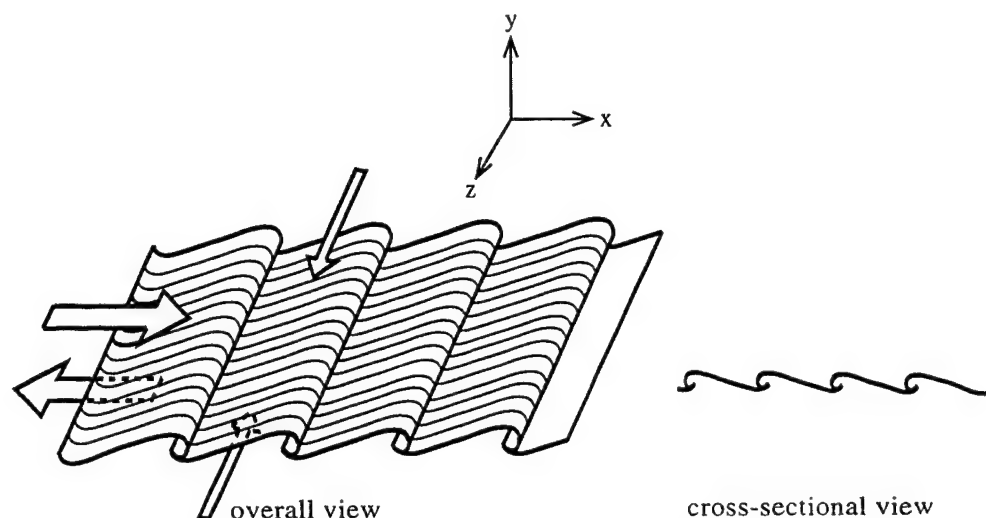


FIGURE 7: A schematic showing the overall picture of a mixing layer with cross shear. A cross section of the streamwise streaks reveals they are "co-rotating" streamwise vortices.

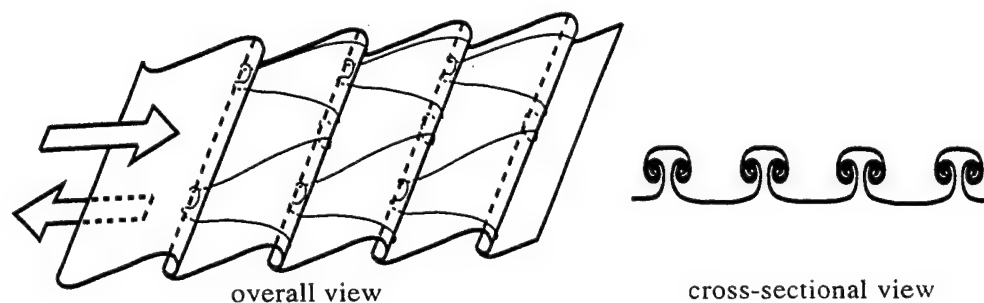


FIGURE 8: A schematic showing the overall picture of a mixing layer with the "mushroom" type streamwise vortices

(in the order of 0.04), making a shear-induced instability likely. Therefore, one cannot think of the flow as two separate and independent arrays of perpendicular Kelvin-Helmholtz vortices superimposed upon each other. Rather the streamwise vortices are a result of a secondary instability in the braid of the primary Kelvin-Helmholtz vortices. Figure 7 shows the overall view of a mixing layer with cross shear. The streamwise vortices appear as streamwise streaks when viewed from top. These streamwise vortices are different from the counter-rotating type as described by Bernal and Roshko (1986). Counter-rotating streamwise vortices would reveal "mushroom" shapes in the x - z plane, like those depicted in figure 8, and the wavelength would not increase because the mushroom structures do not pair.

3.3 Convective instability and localized co-rotating streamwise vortices

The co-rotating streamwise structures could be easily confused with the convective instability that usually presents itself in the form of streamwise rolls of counter-rotating vortices (Thorpe, 1971). The two forms of instabilities are very similar in appearance, being mainly longitudinal and of much smaller wavelengths than that of the primary Kelvin-Helmholtz vortices. However, convective rolls are ruled out in the top view picture of figure 1 because the visualization method is designed to visualize only the interface (see section 2), and any convective instability would reside mainly inside the cores and not be visualized. Moreover, for the experimental parameters represented in figure 1, the wavelength of any convective roll, according to data by Thorpe (1985), would be about 1/3 that of the wavelength of the observed co-rotating streamwise vortices. The flow parameters in most of the present experiments, however, are in the range in which the convective instability would be active. The shadowgraphs of figure 12 of Part I, which visualizes through the whole depth of the fluid, shows an example of convective rolls (marked 'C') which reside in the cores and have roughly the same length scale as the convective rolls observed by Thorpe (1985).

Figure 9 are plan-view shadowgraphs of a two-dimensional stratified mixing layer with $Re=3900$ and $Ri=0.059$, reproduced from Thorpe (1985). It is observed that in the vicinity of a three-dimensional localized pairing, streamwise streaks (marked 'A') of a scale a few times larger than that of the convective rolls (which appears everywhere in the rollers as smaller streamwise streaks) develop in the braid. The amplitude of these streaks, as determined by how sharp the lines appear, tapers away as the distance from the localized pairing increases, indicating that it is intimately tied to the presence of the localized pairing. The scale, behavior and appearance of these structures are very consistent with our observation of co-rotating streamwise vortices. This observation suggests that a mixing layer does not need a "global" cross shear to develop the co-rotating streamwise structures. A three-dimensional phenomenon such as the localized pairing could generate local cross shear induced by the streamwise baroclinic vorticity generation from the high local curvature of the density interface and a large $\partial\rho/\partial z$ component.

3.4 Mixing measurements

The measurements of the mixedness thickness δ_m , defined in section 3 of Part I, show the effect of cross shear on the amount of total mixing attained in the mixing layer. Figure 10 shows the mixedness thickness in mixing layers with different amounts of cross shear, normalized by the mixedness thickness for the no cross shear case. The cross shear is added at about the same time the main Kelvin-Helmholtz instability becomes visible. In figure 11, the amount of cross shear is fixed to 25% of the main shear, but the point in time when the cross shear is introduced to the mixing layer is varied.

Figure 10 illustrates the influence of cross shear upon the total amount of mixing attainable in the mixing layer. In varying the cross shear we keep the total energy budget of the mixing layer fixed, by adjusting the strength of the main shear so that the resultant shear is the same for all cases. With very low cross shear (<15%), the amount of mixing may actually decrease, because taking energy out of the primary shear layer reduces its mixing capacity while the cross shear is not yet strong enough to induce co-rotating streamwise vortices. However, once the cross shear increases past a critical level, the streamwise instabilities are able to form, and a sharp jump

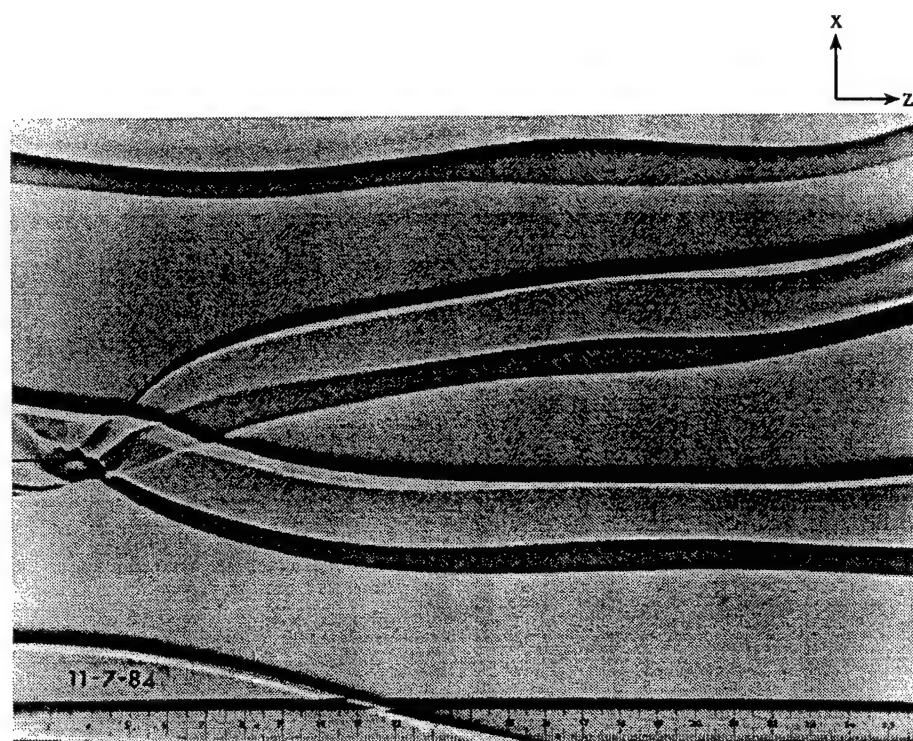
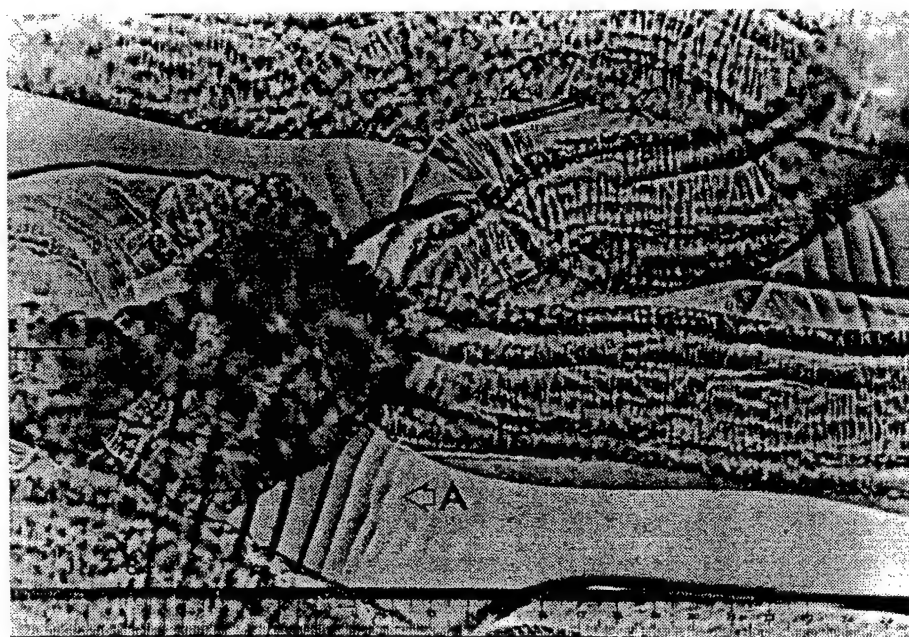
a) $t=3.18$ sec.b) $t=3.47$ sec.

FIGURE 9: Plan-view shadowgraphs from figure 7 in Thorpe (1985) of a two-dimensional stratified mixing layer with $Re=3900$ and $Ri=0.059$ at the onset of Kelvin-Helmholtz instability. The times of the images are from the initiation of tilt. The onset of the main Kelvin-Helmholtz instability (from Thorpe's figure 3) is about 2.9 sec. after tilt. The scale is in cms. (reproduced with the author's permission)

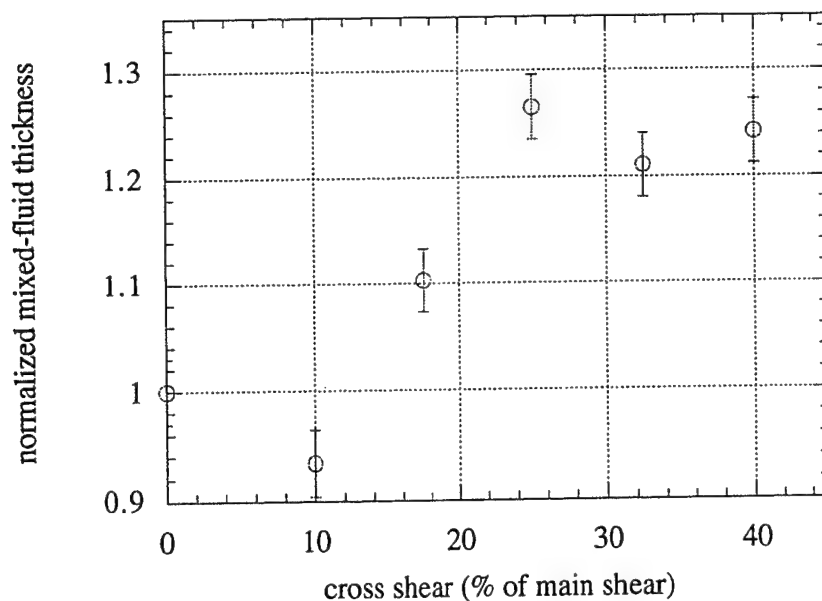


FIGURE 10: Plot of final mixedness thickness vs. cross shear strength. The mixedness thickness is normalized so that the value for the no cross shear case is 1.

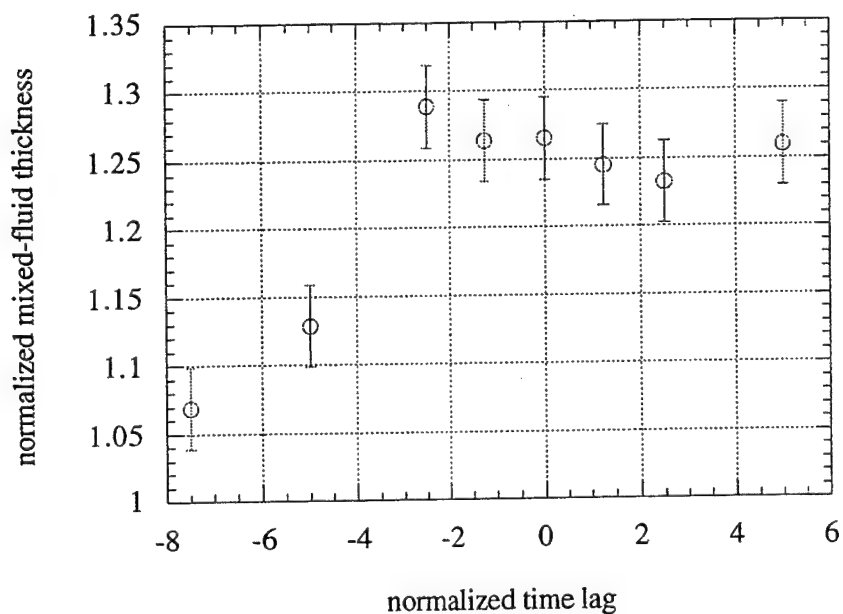


FIGURE 11: Plot of final mixedness thickness vs. the time the cross shear is introduced, for a fixed cross shear strength of 25% of the main shear. Zero time is the point in time when the large-scale Kelvin-Helmholtz instability is first visible. The mixedness is normalized so that the value for the no cross shear case is 1. The time in seconds is normalized with the overturning time scale $\lambda/(\Delta U/2)$.

in mixing efficiency is observed. The co-rotating streamwise vortices, when they are induced, seem to have a considerable effect on the turbulent transports and mixing in the mixing layer. The facts that the co-rotating streamwise vortices reside in the braid and are roughly aligned along the principle axis of strain subject them to large vortex stretching, which means they are highly able to generate a considerable amount of turbulent stresses and small-scaled fluctuations. In a simple two-dimensional mixing layer, most of the entrainment and mixing are observed in the primary vortices. The presence of these streamwise vortices should greatly enhance the rate of entrainment and mixing, as the braid now shares the work of generating small-scale turbulence. Flow visualization also shows that the streamwise vortices affect the flow field around the primary vortices where they interact, possibly increasing the turbulence production and mixing in the primary vortices themselves. Since the streamwise vortices grow through pairings to a size comparable to that of the primary vortices, the interaction between the two perpendicular structures can be expected to be significant, resulting in more mixing and possibly earlier transition as well.

Figure 11 shows the effect of the timing of the cross shear introduction for a fixed cross shear of 25% of the main shear. If the cross shear is introduced too early, the base flow would essentially just respond by reorienting itself into an oblique direction; and the subsequent development of the mixing layer will be no different than if there is only one component of shear, except for some side-wall effects that might create extra sloshing of the fluid. The data show that there is a critical point in time when the introduction of cross shear could create a large jump in mixing. The interpretation is that the cross shear must be introduced after the Kelvin-Helmholtz instability has begun to develop because the presence of the braid is crucial to the generation of the co-rotating streamwise vortices. If the cross shear is introduced too late, the streamwise vortices might not be able to develop so strongly because the stretching in the braid decreases with time as the stable stratification acts to dissipate the rotational energy of the primary vortices by converting it into potential energy in stacked layers of heavy and light fluid.

4. Conclusions

Shear flows in natural settings and practical applications are rarely two-dimensional or steady. When the base flow of a nominally two-dimensional mixing layer changes in such a way as to introduce cross shear to the developing shear layer, a new type of "co-rotating" streamwise vortices may develop in the braid region of the Kelvin-Helmholtz structures. The appearance and dynamical behavior of these co-rotating streamwise vortices are observed to be very similar to those of the primary Kelvin-Helmholtz instability, in that they roll up and pair. The flow field around the braid in a mixing layer with cross shear is dynamically similar to that of a strained vortex sheet, because cross shear introduces a layer of streamwise vorticity and the flow field around the braid induced by the primary Kelvin-Helmholtz structures can be locally estimated as a simple plane strain field. The co-rotating streamwise vortices are found to have a slightly higher wavenumber and a slightly longer normalized time of pairing than the primary Kelvin-Helmholtz instability, in line with the behavior of a strained vortex sheet from the instability and numerical analyses of Lin and Corcos (1984).

Measurements of the density field by the laser-induced fluorescence technique (LIF) show that mixing increases appreciably when a critical level of cross shear is introduced to the mixing layer. It is also found that if cross shear is introduced while

the mixing layer is still essentially a plane two-dimensional shear layer, the mixing layer simply behaves as a normal two-dimensional mixing layer, with the flow redirected along an oblique axis. The co-rotating streamwise vortices and the increased mixing result only when cross shear is introduced after the shear layer starts to roll up and is no longer a plane two-dimensional shear layer.

The behavior of the mixing layer with cross shear must be taken into account when one considers a mixing layer with a complex base flow, such as atmospheric and oceanic shear flows where cross shear is likely to be present. Even in a shear flow where there is no 'global' cross shear, local three-dimensionalities may result in the production of strong localized cross shear. Other scenarios where one might encounter cross shear are stratified mixing layers over three-dimensional topography, wave-mixing layer interaction, and shear flows in complex engineering systems.

We sincerely thank Professors A. Roshko, F. K. Browand, C. D. Winant, P. E. Dimotakis, and A. Leonard for the valuable discussions which contributed a great deal to this work. This work was supported by DARPA grant N00014-91-J-1968.

REFERENCES

- Atsavapranee, P. & Gharib, M. 1994 A plane mixing layer with cross shear. *Physics of Fluids*. **6** (9), 2880.
- Atsavapranee, P. & Gharib, M. 1995 Structures in stratified mixing layers Part 1. The two-dimensional plane mixing layer.
- Bernal, L. P. & Roshko, A. 1986 Streamwise vortex structure in plane mixing layers. *J. Fluid Mech.* **170**, 499.
- Breidenthal, R. 1978 Chemically reacting, turbulent shear layer. Ph.D. thesis, California Institute of Technology.
- Breidenthal, R. 1981 Structure in turbulent mixing layers and wakes using a chemical reaction. *J. Fluid Mech.* **109**, 1.
- Koochesfahani, M. M. & Dimotakis, P. E. 1986 Mixing and chemical reactions in a turbulent liquid mixing layer. *J. Fluid Mech.* **170**, 83.
- Lasheras, J. C. & Choi, H. 1988 Three-dimensional instability of a plane free shear layer: an experimental study of the formation and evolution of streamwise vortices. *J. Fluid Mech.* **189**, 53.
- Lin, S. J. & Corcos, G. M. 1984 The mixing layer: deterministic models of a turbulent flow. Part 3. The effect of plane strain on the dynamics of the streamwise vortices. *J. Fluid Mech.* **141**, 139.
- Martin, J. E. & Meiburg, E. 1991 The three-dimensional evolution of axisymmetric jets perturbed by helical waves. *8th Symposium on Turbulent Shear Flows*, Technical University of Munich, 6-3-1.
- Pierrehumbert, R. T. & Widnall, S. E. 1982 The two- and three-dimensional instabilities of a spatially periodic shear layer. *J. Fluid Mech.* **114**, 59.
- Schowalter D. G. 1993 The effect of stable stratification on three-dimensional structure in shear layers. Ph.D. thesis, University of California at San Diego.
- Thorpe, S. A. 1968 A method of producing a shear flow in a stratified fluid. *J. Fluid Mech.* **32**, 693.
- Thorpe, S. A. 1985 Laboratory observations of secondary structures in Kelvin-Helmholtz billows and consequences for ocean mixing. *Geophys. astrophys. Fluid Dynamics* **34**, 175.

Image correlation velocimetry

P. T. Tokumaru, P. E. Dimotakis

Abstract This paper focuses on the correlation of two successive scalar images for the purpose of measuring imaged fluid motions. A method is presented for deforming, or transforming, one image to another. Taylor series expansions of the Lagrangian displacement field are used, in conjunction with an integral form of the equations of motion, to approximate this transformation. The proposed method locally correlates images for displacements, rotations, deformations, and higher-order displacement gradient fields, and applies a global minimization procedure to insure a global consistency in the results. An integral form of the equations of motion is employed. No explicit spatial or temporal differentiation of the image data is required in estimating the displacement field. As a consequence, this method is appropriate for both continuous-scalar as well as discrete-particle-image data. Successive two-dimensional digital CCD images of fluid motion marked with dye, are used to verify the capabilities of the method. The utility of the method is also illustrated using a pair of Voyager 2 images of Jupiter.

1

Introduction

The application of photographic, CCD, and other forms of imaging for the purpose of estimating flow velocities, has been investigated by many researchers in fields ranging from fluid mechanics to vision research. In the most common methods for measuring fluid-flow velocities, the fluid is seeded with particles, or markers, that can be easily imaged and tracked. An extensive review of methods using particle and speckle images for fluid flow measurement is presented by Adrian (1991). Brandt & Hesselink (1990) apply a three-dimensional, linear, auto-correlation method to examine the structure of smoke flow (imaged as a con-

tinuous scalar) over a delta wing. Their method uses a three-dimensional data set comprised of two spatial-dimensions and time, or three spatial dimensions, and is used for detecting strong vortical motions. The estimation of the motion and deformation of solids is closely related to that of fluids. A method of determining displacements and stress intensity factors in solids, using white light speckle images and image correlation techniques, is presented in McNeill et al. (1987). In the absence of particles, flows have also been tagged with a line or grid, e.g., using laser-induced photochemical reactions (Falco and Chu 1987), or laser-induced fluorescence (Miles et al. 1989). When this is not possible, one can use markers that occur naturally in the flow, e.g., Bindschadler and Scambos (1991) have correlated the translation of distinct surface features in ice flows to determine flow velocities. In a method similar to that pioneered by Horn and Schunck (1981) and Hildreth (1984), for determining "optical flow" (see Sects. 1.1 and 1.3), Dahm et al. (1992) have claimed that the scalar transport equation can be applied to recover a simulated velocity field.

Determining motion from successive images is also of interest in animation, as well as the study of biological and robotic vision. Most investigations along these lines have taken the form of extracting the motion of objects in an image and, as a consequence, they focus on the motion of rigid objects and their representations. See Hildreth and Koch (1987) for a review and Murray and Buxton (1990). This approach is somewhat different from the interests of Fluid Mechanics where the object of interest is a fluid, highly deformable and often compressible. Nevertheless, many results from object motion research apply directly to the motion of fluids and solids.

The proposed Image Correlation Velocimetry (ICV) method that will be developed in the discussion that follows borrows from the correlation methods used in measuring fluid flow and the deformation of solids, outlined in Sect. 1.2, and the gradient methods used in measuring optical and fluid flows, outlined in Sect. 1.3. The proposed new nomenclature is justified, in our opinion, because most image velocimetry methods documented to date are specific to the particular type of image data, e.g., Particle Imaging Velocimetry (Adrian 1991, Willert and Gharib 1991), Scalar Image Velocimetry (Dahm et al. 1992), and Particle Image Distortion (Huang et al. 1993b). In contrast, because of its hybrid nature and the fact that it does not rely on temporal or spatial derivative information to be estimated from the image sequence, the proposed Image Correlation Method can be employed to analyze a wide range of image data for the convecting motion field, e.g., laboratory fluid flow images employing either particle or dye markers;

Received: 12 January 1993 / Accepted: 3 December 1993

P. T. Tokumaru, P. E. Dimotakis
Graduate Aeronautical Laboratories,
California Institute of Technology 301-46,
Pasadena, CA 91125, USA

Correspondence to: P. E. Dimotakis

We would like to acknowledge the support of this effort by AFOSR Grant Nos. 90-0304 and F49620-92-J-0290, and a joint contract with M. Gharib by DARPA/Navy Grant No. N00014-91-J-1968. We would also like to acknowledge the many contributions by Dan Lang in the development and integration of the image-data acquisition systems employed in these investigations. Finally, we would like to thank Stewart (Andy) Collins and the Detector Advanced Development group of the Imaging Systems Section, at JPL, for their collaboration.

direct images of the flowing fluid itself, as in Rayleigh-scattering images in gas-phase flows; satellite images of cloud or glacial motion; etc.

1.1

Fluid flow and image flow

If we imagine an image sequence as a function of time, we can describe the associated evolution as a succession of mappings that produce each image from the previous one. For each neighborhood within an image at time t_1 , we can express the mapping in terms of the local displacement, ξ , of the neighborhood at \mathbf{x} , that produces the corresponding neighborhood in the next image at time t_2 in the sequence. We can define the associated image-flow velocity field, or optical-flow velocity field, as referred to by Horn and Schunck (1981), Hildreth (1984), Hutchinson et al. (1988), and Murray and Buxton (1990), i.e., $\mathbf{u}(\mathbf{x}, t)$, as the time-rate-of-change of the mapping displacement field, $\xi(\mathbf{x}, t)$.

Considering a sequence of (Lagrangian) markers that are convected by a fluid, we may ask for the succession of mappings that transform the coordinates of the markers at time t_1 , to the corresponding coordinates at time t_2 . These mappings may also be described in terms of displacement fields, $\xi_r(\mathbf{x}, t)$, of the fluid elements, however. We associate the time-rate-of-change of the associated fluid-displacement field, $\xi_r(\mathbf{x}, t)$ with the fluid-flow velocity field, $\mathbf{u}_r(\mathbf{x}, t)$.

In this discussion, we will propose a method of extracting the image-flow velocity, $\mathbf{u}(\mathbf{x}, t)$. We recognize that the extent to which this represents the fluid-flow velocity field, $\mathbf{u}_r(\mathbf{x}, t)$, is an important but separate problem that can only be discussed in the context of the particular fluid-imaging application in each case.

By way of example, image-flow may significantly depart from fluid-flow. Perspective may also play a role in generating an optical flow which is not coincident with the true flow, e.g., if an observer moves toward an object, there will be an outward flow associated with the apparent increase in size of the object within an image (Gibson 1950). As is illustrated by the image sequence of a shadow moving across the ground, a shadow can have a perceived velocity that is markedly different from that of the object casting the shadow, or the (stationary) ground. Similarly, a rotating, featureless sphere will have no apparent velocity (Hutchinson et al. 1988).

A related problem is also illustrated by the difficulty in both laser-Doppler and particle-image velocimetry, wherein the measurement method relies on particle displacement and velocity fields, which may or may not adequately track the fluid-velocity fields of interest. Additional problems must be confronted when the motion is of a continuous scalar, the most prominent being the so-called "aperture" problem to be discussed below in the context of gradient methods (Sect. 1.3). Additional discussions of the aperture problem and other issues associated with interpreting image flow can be found in the references cited in the first paragraph of this section. A separate issue arises when the dimensionality of the image information does not match that of the fluid flow from which it is derived. We will also discuss this issue below, in the context of two-dimensional scalar images of three-dimensional fluid-flow (Sect. 4).

This paper presents the foundations of the Image Correlation Velocimetry (ICV) method along with a discussion of three examples of image/fluid flows. In the first example, of a randomly distributed dye marker in glycerine, the fluid flow is two-dimensional, as is the image, and the image-flow is expected to be a good representation of the fluid-flow. In the second example, of planar laser-induced fluorescence images of the flow past a cylinder, both the flow velocity and the scalar (fluorescent dye) distribution fields are three-dimensional, while the image is two-dimensional. In this case, the image-flow field need not necessarily represent the fluid-flow velocity field. Finally, the motion of imaged quantities in the Jovian atmosphere do not follow any simple equation of motion. As a result, the derived image-flow velocity field can be expected to vary from that of the fluid flow, though the image flow is interesting in itself.

1.2

Correlation methods

Several techniques for determining (image) flow velocities from particle image pairs (e.g., Willert and Gharib 1991) employ an optimization that relies on some form of a cross-correlation function, e.g.,

$$\max_{\mathbf{a}} \int_A E_0(\mathbf{x}) E_1(\xi) d^2\mathbf{x}, \quad (1)$$

with

$$\xi = \mathbf{x} + \mathbf{a}, \quad (2)$$

where \mathbf{a} is a vector parameter to be determined by the optimization procedure and A is the correlation region. The distribution of the image irradiance, $E(\mathbf{x}, t)$, is known at times t_0 and t_1 , i.e.,

$$E_0(\mathbf{x}) \equiv E(\mathbf{x}, t_0) \quad \text{and} \quad E_1(\mathbf{x}) \equiv E(\mathbf{x}, t_1). \quad (3)$$

The average velocity, \mathbf{u}_A , within the correlation region is then approximated by,

$$\mathbf{u}_A \approx \frac{\mathbf{a}}{t_1 - t_0}. \quad (4)$$

The drawback of this method, having only two parameters to quantify the motion, is that it cannot resolve displacements properly where there are large displacement gradients within the correlation region. Anticipating this problem, and being very interested in displacement gradients, researchers in Solid Mechanics apply techniques which include higher-order deformations of the displacement field within a correlation volume. For example, McNeill et al. (1987) describe a method whereby a model of the image displacement field (mapping) is used in a least squares optimization procedure, i.e.,

$$\min_{\mathbf{a}, \nabla \mathbf{a}} \int_A [E_0(\mathbf{x}) - E_1(\xi)]^2 d^2\mathbf{x}. \quad (5)$$

The affine mapping,

$$\xi = \mathbf{x} + \mathbf{a} + (\nabla \mathbf{a}) \cdot \mathbf{dx}, \quad (6)$$

is used as an example of such a function. The displacement \mathbf{a} and the four components of $\nabla \mathbf{a}$ are treated as parameters to be determined by the optimization procedure. However, any physically motivated mapping can be used in place of Eq. (6).

More recently, Huang et al. (1993a, b) have introduced a (similar) Particle Image Distortion (PID) technique for images of fluid flow. For a first approximation of the velocity and velocity gradient fields, the PID method employs conventional Particle Image Velocimetry (PIV). Subsequent approximations for the velocity and velocity gradient fields include a locally linear image deformation, described by Eq. (6), and an optimization procedure for the parameters \mathbf{a} and $\nabla \mathbf{a}$.

In both these methods, the image data are integrated over a region and require no spatial differentiation. Since, for two- and three-dimensional images, only a few parameters are extracted from the optimization, these methods are relatively immune to noise and lend themselves to fast solutions.

1.3

Gradient methods

A method for determining optical-flow velocities was presented by Horn and Schunck (1981). The fundamental equation used by Horn and Schunck to determine the optical flow was,

$$\frac{\partial E}{\partial t} + \mathbf{u} \cdot \nabla E = 0, \quad (7)$$

where E is the image irradiance (a scalar) and \mathbf{u} is the optical velocity. The differential terms, $\partial E / \partial t$ and ∇E , can be estimated from the image data and the component of \mathbf{u} along ∇E is calculated using eq. (7). Methods employing equations of this type are called gradient schemes.

The idea of including constraints in the optimization process that determines the velocity field, over an area, is important in the context of the present method. We note that Horn and Schunck (1981) applied a global constraint to Eq. (7) (in two dimensions), by solving for $\mathbf{u}(\mathbf{x}, t)$ using an optimization, i.e.,

$$\min_{\mathbf{u}(\mathbf{x}, t)} \int_A \left(\left[\frac{\partial E}{\partial t} + \mathbf{u} \cdot \nabla E \right]^2 + \sigma^2 \right) d^2 \mathbf{x}, \quad (8)$$

where σ represents the constraint cost function in the optimization process. In particular, Horn and Schunck chose smoothness as a constraint, i.e.,

$$\sigma^2 = \sum_{i,j=1,2} \gamma \left(\frac{\partial u_i}{\partial x_j} \right)^2, \quad (9)$$

where γ balances the relative cost of smoothness and of Eq. (7). Numerous refinements of, and additions to, this method can be found in Hildreth (1984). Note that the constraint cost function in Eq. (8), σ , need not be included in the optimization integral. Instead, a constraint could be included as a feature of the optimization technique (Murray and Buxton 1990).

Several problems appear when considering gradient methods. The most prominent is that no unique velocity can be calculated using Eq. (7), if there are no features or gradients in the image, i.e., if $\nabla E = 0$. See for an example in Sect. 1.1. In addition, because Eq. (7) involves only the component of \mathbf{u} along ∇E , velocity components along isointensity contours of E cannot be determined using Eq. (7) alone. This limitation was designated as the "aperture problem" by Wallach (1976). See also Hildreth (1984) and Hutchinson et al. (1988). The terminology is somewhat misleading and only used here in reference to the convention.

The uncertainty of the so-called aperture problem can be resolved in some cases by applying constraints to the motion, e.g., the motion is of a rigid body (see Murray and Buxton 1990, for example), or a limited class of deformable bodies (see Terzopoulos and Metaxas 1991). Pearlstein and Carpenter (1993) have argued that the so-called aperture problem can also be resolved if the flow field is two-dimensional and solenoidal, or if the flow field is three-dimensional and estimated from two uncorrelated scalars, convected by the same flow.

Gradient schemes also have the problem that spatial and temporal differentiation of noisy image data is necessary. Another problem in differentiating the image data is related to the Nyquist sampling criterion: aliasing in the image-intensity data can affect velocity estimates. To minimize this problem in taking the gradient, the motion between images should be less than half the smallest local spatial scale, λ_E , of the E -field, i.e.,

$$\frac{|\mathbf{u}|(t_1 - t_0)}{\lambda_E} < \frac{1}{2} \quad (10)$$

(cf. Eq. (7)), where $(t_1 - t_0)$ is the time between images. The signal-to-noise requirements, along with those restrictions inherent in Eq. (10), can lead to prohibitive data acquisition requirements, e.g., for a given flow velocity, $|\mathbf{u}|$, a small λ_E requires a high spatial imaging resolution and dictates a fast data acquisition system, i.e., $(t_1 - t_0)$.

In the context of fluid-flow measurements derived from consecutive images of a scalar in turbulent flows, Dahm et al. (1992) propose to estimate the velocity field using a method similar to those described in Hildreth (1984). Dahm et al. state that the so-called aperture problem does not apply in cases where the spatial scale of the E -field can be assumed to be sufficiently smaller than that of the \mathbf{u} -field. Assuming that the so-called aperture problem does not exist in their flows, Dahm et al. proceed by estimating several velocity vectors (using Eq. (7)) along different scalar gradient directions within a small region of the flow. As a first approximation in an iterative procedure, they assume a locally-uniform flow field. Then a velocity, consistent with this assumption and the estimated velocity components along the scalar gradient directions, is calculated. Subsequent approximations are attempted, iteratively, by allowing local velocity gradients. This restriction in the form of the permitted (local) velocity field is similar to those applied in the correlation methods of Sect. 1.2, i.e., Eqs. (4) and (6).

2

Proposed methodology

A succession of images can represent anything from the motion of cars on a highway to the transport of a dye marker in water. As discussed in Sect. 1.1, we take the view that given successive image representations, there exists a transformation, or mapping, of the local image intensity data that takes one image to the next. In many cases, while the equation of motion of the imaged field may be known, the mapping taking one image to another may not be. Successive images combined with the equations of motion, however, often allow us to approximate the mapping.

The mapping of one image to the next can be developed by considering the Lagrangian displacement field $\xi(\mathbf{x}, t)$ of the image sequence. Specifically, if \mathbf{x} is the coordinate of a point on an image at some initial time t_0 , $\xi(\mathbf{x}, t)$ represents the coordinate of this point in a subsequent image recorded at a later time t . If we imagine the image sequence as the result of a continuous recording process, we can assign a Lagrangian image-flow velocity field, also referred to as "optical flow" in the discussion and literature cited above, i.e.,

$$\mathbf{u}[\xi(\mathbf{x}, t), t] = \frac{d}{dt} \xi(\mathbf{x}, t), \quad (11)$$

to the continuous displacement field $\xi(\mathbf{x}, t)$ that takes an initial point \mathbf{x} in the image recorded at time t_0 , to the point $\xi(\mathbf{x}, t)$ on the image recorded at time t .

In the proposed implementation, local series approximations for the displacement mapping are used in conjunction with an integral form of the equations of motion. A global nonlinear correlation (optimization) process is employed to estimate the image-flow velocity, vorticity, deformation rate, etc., of the imaged data field. "Series", in this discussion, will denote "Taylor series."

In the context of fluid mechanics measurements, we will focus on images of continuous passive, convected scalars, e.g., dye markers, carried by a fluid. As will be illustrated using a pair of Voyager 2 images of Jupiter (Sect. 5), however, any marker in the flow can be used.

The method will be developed for three spatial dimensions and time, and can yield three-dimensional velocity fields. The method can also obtain two-dimensional velocity fields from images of two-dimensional flows. In a concession to the limitations of typical data acquisition systems today, however, the method will be applied here to a two-dimensional flow and also to two-dimensional slices of three-dimensional flows. A short discussion of the implications of correlating two-dimensional slices of three-dimensional data is presented in Sect. 4. In some cases, two-dimensional imaging devices can be used to obtain approximations to three-dimensional image data, developing with time, e.g., Dahm et al. (1991, 1992).

2.1

Fluid displacement and equations of motion

To see how the image-flow velocity field can be calculated from three-dimensional image data sets, spaced in time, first consider a Lagrangian description of a flow being imaged. In what follows, we will assume that the three-dimensional displacement-field, $\xi(\mathbf{x}, t)$, also represents the fluid-flow displacement field. Figure 1 illustrates the motion of fluid particles within a volume, V . Fluid elements at \mathbf{x} , in a neighborhood V , at time t_0 , are convected to locations $\xi(\mathbf{x}, t)$ at a later time t . The displacement field, $\xi(\mathbf{x}, t)$, can be thought of as a transformation of the field \mathbf{x} , at time t_0 , to $\xi(\mathbf{x}, t)$.

Using this Lagrangian field, $\xi(\mathbf{x}, t)$, one could, in principle, integrate the equation of motion of the imaged scalar, i.e.,

$$\frac{\partial c}{\partial t} + \mathbf{u}_f \cdot \nabla c = \mathcal{D} \nabla^2 c, \quad (12)$$

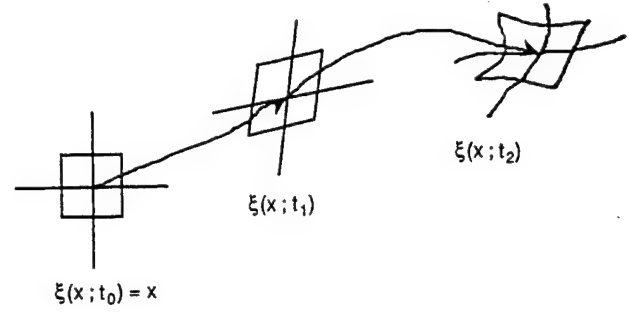


Fig. 1. Motion of a fluid volume.

to obtain

$$c_1[\xi(\mathbf{x}, t_1)] - c_0[\xi(\mathbf{x}, t_0)] - \mathcal{D} \int_{t_0}^{t_1} \nabla^2 c[\xi(\mathbf{x}, \tau), \tau] d\tau = 0, \quad (13)$$

where

$$c_0(\mathbf{x}) \equiv c(\mathbf{x}, t_0) \quad \text{and} \quad c_1(\mathbf{x}) \equiv c(\mathbf{x}, t_1) \quad (14)$$

represent the $c(\mathbf{x}, t)$ -field at times t_0 and t_1 , respectively, \mathcal{D} is the appropriate diffusion coefficient, and \mathbf{u}_f is the fluid velocity.

In the first two examples, the motion of food coloring in glycerine (Sect. 3) and dilute fluorescein in water (Sect. 4), are examined. In these flows, the diffusion of the dye markers in the time interval between successive images, is relatively small and may be neglected, i.e., for these data,

$$\frac{(t_1 - t_0)\mathcal{D}}{\ell^2} \ll 1, \quad (15)$$

where, $t_1 - t_0 = 0.1$ s is the time between images, $\ell = 50$ to $80 \mu\text{m}$ is the imaging resolution, and the diffusion coefficients are no larger than $\mathcal{D} = 10^{-9} \text{ m}^2/\text{s}$. In addition, we note that the Schmidt number is large in both flows, i.e.,

$$Sc \equiv v/\mathcal{D} > 10^3, \quad (16)$$

where v is the kinematic viscosity.

In cases where the diffusion of the imaged scalar can be ignored, Eq. (13) becomes

$$c_1[\xi(\mathbf{x}, t_1)] - c_0[\xi(\mathbf{x}, t_0)] = 0. \quad (17)$$

Equation (17) represents a significant simplification over Eq. (12), its differential counterpart. It contains no spatial, or temporal, derivatives and as such is less restrictive than gradient methods discussed earlier (Sect. 1.3).

Using the integral equation of motion (Eq. (17)) in place of the differential equation of motion (Eq. (7)) in the optimization (Eq. (8)), and generalizing the optimization to three dimensions then yields an expression for determining $\xi(\mathbf{x}, t_0)$ and $\xi(\mathbf{x}, t_1)$, i.e.,

$$\min_{\xi(\mathbf{x}, t_0), \xi(\mathbf{x}, t_1)} \int_V \left(\left\{ c_1[\xi(\mathbf{x}, t_1)] - c_0[\xi(\mathbf{x}, t_0)] \right\}^2 + \sigma^2 \right) d^3\mathbf{x}, \quad (18)$$

corresponding to an appropriate choice of the smoothing function, σ^2 .

In the spirit of the correlation methods discussed in Sect. 1.2, where the type of motion within the correlation volume is limited to translation alone, the present method

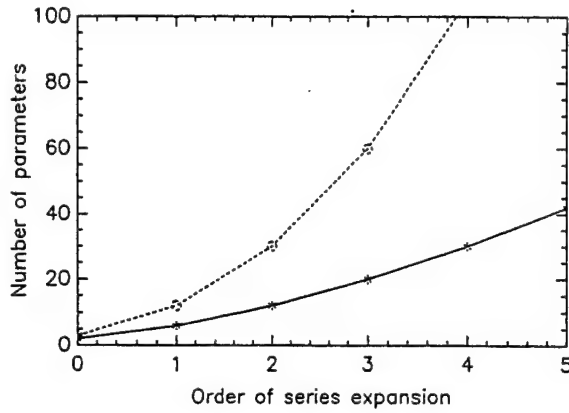


Fig. 2. This plot shows the rapid increase in the number of parameters used in the optimization procedure with increasing order of the series expansion. "x" is for a 2-D expansion and "o" for 3-D.

restricts $\xi(\mathbf{x}, t_0)$ and $\xi(\mathbf{x}, t_1)$ to a truncated series approximation for $\xi(\mathbf{x}, t)$. While many representations of the displacement field could be employed, a (Taylor) series representation is used because the first two orders in the series expansion correspond to physical fluid mechanical quantities, i.e., the velocity vector and the velocity gradient tensor. More importantly, the series approximation has the additional benefit of enforcing smoothness in the displacement and displacement gradient fields within a correlation volume. In addition, if the spatial scale of the E -field is everywhere sufficiently smaller than that of the $\nabla \mathbf{u}$ -field, then the solution can be unique.

2.2

Displacement field and kinematic quantities

In the case of images of fluid-flow, the quantity $\xi(\mathbf{x}, t)$ is a complicated non-linear function of the imaging process, the nonlinear convection dynamics, and \mathbf{x} . Local estimates of this function can be made by Taylor series, expanding $\xi(\mathbf{x}, t)$ in space, at some time t , in an image correlation volume, V . This yields,

$$\begin{aligned} \xi(\mathbf{x}, t) = & \xi(\mathbf{x}_c, t) + (\mathbf{x} - \mathbf{x}_c) \cdot \nabla \xi(\mathbf{x}_c, t) \\ & + \frac{1}{2!} [(\mathbf{x} - \mathbf{x}_c) \cdot \nabla]^2 \xi(\mathbf{x}_c, t) + \dots \end{aligned} \quad (19)$$

In this expression, \mathbf{x}_c denotes the center of the image correlation volume, V , at time t_0 , and $\nabla \xi(\mathbf{x}_c, t)$ denotes the gradient of $\xi(\mathbf{x}, t)$ with respect to \mathbf{x} , evaluated at \mathbf{x}_c . Figure 2 plots the number of parameters used in the optimization process as a function of the order used in the series expansion, for two and three dimensions. Figure 3 illustrates the effect of the various orders of the expansion on a two-dimensional square "volume."

Using a finite difference approximation in time for the velocity, Eq. (11), and the series representation, Eq. (19), evaluated at times t_0 and t_1 , yields an estimate for the velocity within the correlation volume, i.e.,

$$\mathbf{u}[\xi(\mathbf{x}, t), t] \simeq \frac{\xi(\mathbf{x}, t_1) - \xi(\mathbf{x}, t_0)}{t_1 - t_0} \quad (20)$$

Translation (no deformation)

Linear deformation

Quadratic deformation

Cubic deformation

Combined translation and deformation

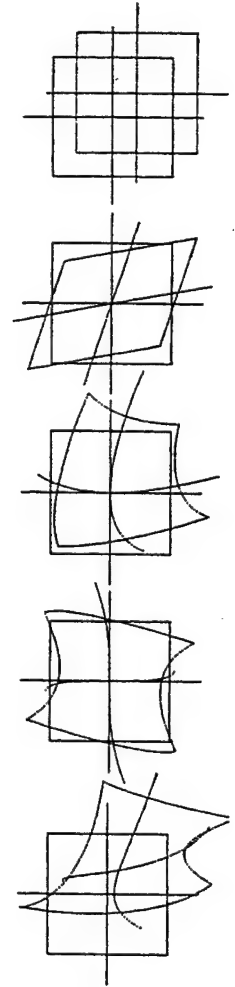


Fig. 3. The effect of translation and various orders of deformation on a two-dimensional square "volume." See Sect. 2.2.

where $t_0 \leq t \leq t_1$. Similarly, taking the spatial gradient of Eq. (20) yields an expression for the velocity gradient tensor within the correlation volume, i.e.,

$$\nabla \mathbf{u}[\xi(\mathbf{x}, t), t] \equiv \frac{\partial}{\partial \mathbf{x}} \mathbf{u}[\xi(\mathbf{x}, t), t] \simeq \frac{\nabla \xi(\mathbf{x}, t_1) - \nabla \xi(\mathbf{x}, t_0)}{t_1 - t_0} \quad (21)$$

Vorticity, divergence, and strain rate can then be obtained from the components of the estimated velocity gradient tensor, Eq. (21).

There is some freedom in choosing the coordinate transformation at the initial time t_0 , $\xi(\mathbf{x}, t_0)$. Our choice is to have the coordinate description at the initial time t_0 correspond with the local Eulerian coordinates at that time, i.e.,

$$\xi(\mathbf{x}, t_0) = \mathbf{x}. \quad (22)$$

In terms of the series expansion, Eq. (19), this means that

$$\nabla \xi(\mathbf{x}_c, t_0) \equiv \mathbf{I}, \quad (23)$$

where \mathbf{I} is the identity tensor, and all other higher order derivative terms are identically zero.

2.3

Seeking a global solution

A solution for the coefficients of the series expansion, i.e., $\xi(x, t)$, $\nabla \xi(x, t)$, etc., can be obtained in a neighborhood around x_c using the expansions for $\xi(x, t_0)$ and $\xi(x, t)$ from Sect. 2.2. The unknown coefficients of the series expansion, $\xi(x_c, t)$, $\nabla \xi(x_c, t)$, etc., are treated as parameters in an optimization process. To minimize the difference between two data sets (images), in a least-squares sense, for a single correlation volume, we use the optimization, Eq. (18), in conjunction with the series approximations developed in Sect. 2.2, i.e.,

$$\min_{\xi(x_c, t), \nabla \xi(x_c, t), \dots} \int_V \{ \{ c_1 [\xi(x_c, t_1) + (x - x_c) \cdot \nabla \xi(x_c, t_1) + \dots] - c_0 [x] \}^2 + \sigma^2 \} d^3x. \quad (24)$$

The optimization implied in Eq. (24), combines many of the good features of correlation and gradient methods, while eliminating many of the deficiencies. Specifically, this optimization method has high immunity to noise, uses the equations of motion, can incorporate constraints, requires no differentiation to calculate the displacement field, and can capture displacement gradients within a correlation volume.

In principle, a single correlation volume covering the entire image and a series approximation of a high enough order can be used to capture the entire image displacement field. In practice, however, employing a series approximation beyond quartic terms is impractical because of the rapid increase in the number of parameters in the optimization process with increasing order (see Fig. 2), and the associated increase in the computational time and complexity. In the present calculations, when the quartic terms are not sufficient to capture the image deformation over the entire flow field using a single volume, as is usually the case, several series expansions residing in smaller, adjacent, correlation volumes are used in place of the single large volume.

To construct a global optimization using a number of local series expansions, we require that neighboring correlation volumes, with independent series expansions, must yield consistent results. In the present method, we use the expansion for the displacement field about one correlation volume to estimate those of its neighbors. The displacement field of these neighbors is also estimated in terms of their own series expansions. The root-mean-square difference between displacements, displacement gradients, and higher-order gradients, estimated by neighboring correlation volumes is applied as a constraint cost function, σ^2 . Since it is necessary to refer to a number of series expansions, it is useful to define the Taylor series for $\xi(x, t)$ in a neighborhood V_i centered about x_c , as

$$\xi_i(x, t) \equiv \xi(x_c, t) + (x - x_c) \cdot \nabla \xi(x_c, t) + \frac{1}{2!} [(x - x_c) \cdot \nabla]^2 \xi(x_c, t) + \dots \quad (25)$$

When series expansions about multiple points are then employed, the minimization, Eq. (24), is modified, i.e.,

$$\min_{\xi(x_c, t), \nabla \xi(x_c, t), \dots} \int_V \{ \{ c_1 [\xi_i(x, t)] - c_0 [x] \}^2 + \sigma_i^2 \} d^3x, \quad \forall i. \quad (26)$$

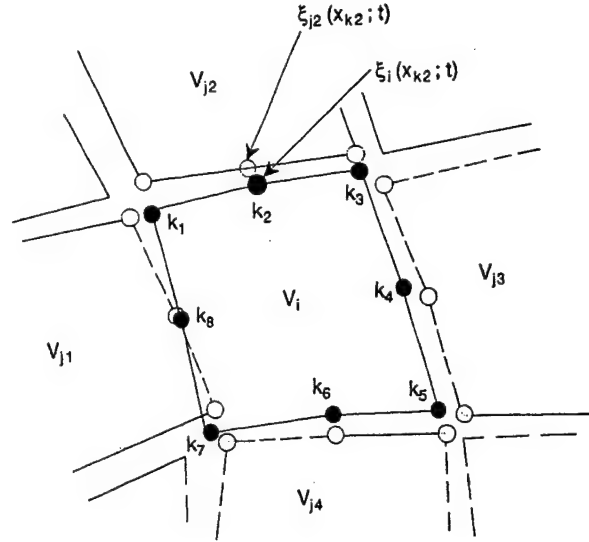


Fig. 4. The k_1 - k_8 solid circles denote the points in V_i used by the constraint cost function σ_i . The empty circles denote their counterparts estimated by the neighbors V_{j1} - V_{j4} .

This minimization is performed within all the V_i , simultaneously, and the constraint cost function,

$$\sigma_i^2 = \sum_k [\gamma_0 | \xi_i(x_k, t) - \xi_{j(k)}(x_k, t) |^2 + \gamma_1 | \nabla \xi_i(x_k, t) - \nabla \xi_{j(k)}(x_k, t) |^2 + \gamma_2 | \nabla^2 \xi_i(x_k, t) - \nabla^2 \xi_{j(k)}(x_k, t) |^2 + \dots] \quad (27)$$

is applied to provide global continuity of the solution. ξ_j denotes the series expansion about the " $j(k)$ " neighbors of V_i , x_k denotes the " k " points of comparison between the solutions ξ_i and ξ_j , and the γ 's are weighting constants. See Fig. 4. In the figure, eight points about each correlation volume are used for comparison, three with each of four neighbors. In practice, we use 15 points per neighbor.

The present implementation of the method solves Eq. (26) for two-dimensional flow up to quartic terms in the local series expansions. The optimization of Eq. (24) is accomplished using a multidimensional minimization process, with image data between pixels estimated using bilinear interpolation. At the local level, minimization is performed using modified versions of both Powell's Method for multi-dimensions, or the Continuous Minimization by Simulated Annealing routines found in Press et al. (1992). These routines are integrated into an iterative global minimization routine.

2.4

Minimization parameters in two dimensions

Typical CCD imaging technologies today are limited to two-dimensional (spatial) data. This is not a problem if the flow being imaged is also two-dimensional. This section describes how the method is applied in two dimensions. First, we develop the terms of the series expansion, Eq. (19), for two-dimensional flow. With the two-dimensional vector

$$\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \equiv x - x_c, \quad (28)$$

as the position, \mathbf{x} , relative to the center of the correlation volume, \mathbf{x}_c , the terms of the series expansion at a time t_1 appear as a constant term,

$$\xi(\mathbf{x}_c, t_1) = \begin{bmatrix} \alpha_1 \\ \alpha_0 \end{bmatrix}, \quad (29)$$

where the α_i are the vector coordinates of the center of the correlation volume at the time t_1 , i.e.,

$$\alpha_i \equiv \xi_i(\mathbf{x}_c, t_1), \quad (30)$$

a linear term,

$$(\mathbf{x} - \mathbf{x}_c) \cdot \nabla \xi(\mathbf{x}_c, t_1) = \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} \\ \alpha_{2,1} & \alpha_{2,2} \end{bmatrix} \cdot \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}, \quad (31)$$

where the $\alpha_{i,j}$ represent the first order deformations and rotations of the image field within the correlation volume, i.e.,

$$\alpha_{i,j} \equiv \frac{\partial \xi_i(\mathbf{x}_c, t_1)}{\partial x_j}, \quad (32)$$

a quadratic term,

$$\frac{1}{2!} [(\mathbf{x} - \mathbf{x}_c) \cdot \nabla]^2 \xi(\mathbf{x}_c, t_1) = \begin{bmatrix} \alpha_{1,11} & \alpha_{1,12} & \alpha_{1,22} \\ \alpha_{2,11} & \alpha_{2,12} & \alpha_{2,22} \end{bmatrix} \cdot \begin{bmatrix} \delta_1^2 \\ \delta_1 \delta_2 \\ \delta_2^2 \end{bmatrix}, \quad (33)$$

where

$$\alpha_{i,jk} \equiv \frac{1}{2!} \left(\frac{\partial^2 \xi_i(\mathbf{x}_c, t_1)}{\partial x_j \partial x_k} \right), \quad (34)$$

a cubic term,

$$\frac{1}{3!} [(\mathbf{x} - \mathbf{x}_c) \cdot \nabla]^3 \xi(\mathbf{x}_c, t_1) = \begin{bmatrix} \alpha_{1,111} & \alpha_{1,112} & \alpha_{1,122} & \alpha_{1,222} \\ \alpha_{2,111} & \alpha_{2,112} & \alpha_{2,122} & \alpha_{2,222} \end{bmatrix} \cdot \begin{bmatrix} \delta_1^3 \\ \delta_1^2 \delta_2 \\ \delta_1 \delta_2^2 \\ \delta_2^3 \end{bmatrix}, \quad (35)$$

with

$$\alpha_{i,jkl} \equiv \frac{1}{3!} \left(\frac{\partial^3 \xi_i(\mathbf{x}_c, t_1)}{\partial x_j \partial x_k \partial x_l} \right), \quad (36)$$

and so on for higher order terms. The $\alpha_{i,jk}$ and $\alpha_{i,jkl}$ are, respectively, related to the second and third derivatives of the displacement field within the correlation volume, i.e., by Eqs. (34) and (36).

The velocity and velocity gradient (Eqs. (20) and (21)) can also be written in terms of the parameters of Eqs. (29) and (31) and the series expansions at times t_0 and t_1 , i.e.,

$$\mathbf{u}(\mathbf{x}_c, t) \equiv \begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{t_1 - t_0} \begin{bmatrix} \alpha_1 - x_c \\ \alpha_2 - y_c \end{bmatrix}, \quad (37)$$

where

$$\mathbf{x}_c = \begin{bmatrix} x_c \\ y_c \end{bmatrix} \quad (38)$$

and

$$\nabla \mathbf{u}(\mathbf{x}_c, t) \equiv \begin{bmatrix} \partial u / \partial x & \partial u / \partial y \\ \partial v / \partial x & \partial v / \partial y \end{bmatrix} = \frac{1}{t_1 - t_0} \begin{bmatrix} \alpha_{1,1} - 1 & \alpha_{1,2} \\ \alpha_{2,1} & \alpha_{2,2} - 1 \end{bmatrix}. \quad (39)$$

Alternatively, the velocity gradient can be written in terms of the in-plane vorticity and rate-of-strain tensor, i.e.,

$$\nabla \mathbf{u}(\mathbf{x}_c, t) = \begin{bmatrix} 0 & -\omega_z/2 \\ \omega_z/2 & 0 \end{bmatrix} + \begin{bmatrix} s_{xx} & s_{xy} \\ s_{xy} & s_{yy} \end{bmatrix}, \quad (40)$$

where ω_z is the vorticity, i.e.,

$$\omega_z = \frac{\alpha_{2,1} - \alpha_{1,2}}{t_1 - t_0}, \quad (41)$$

and s_{xx} , s_{yy} , and s_{xy} , are the components of the rate-of-strain tensor, i.e.,

$$s_{xx} = \frac{\alpha_{1,1} - 1}{t_1 - t_0}, \quad (42)$$

$$s_{yy} = \frac{\alpha_{2,2} - 1}{t_1 - t_0}, \quad (43)$$

and

$$s_{xy} = \frac{1}{2} \frac{\alpha_{1,2} + \alpha_{2,1}}{t_1 - t_0}. \quad (44)$$

An interesting quantity to consider is the second invariant of the rate-of-strain tensor (see Cantwell 1992, for example), i.e.,

$$-2q_s \equiv s_{xx}^2 + 2s_{xy}^2 + s_{yy}^2. \quad (45)$$

3

Couette flow between concentric cylinders

An apparatus to generate a Couette flow between concentric cylinders was fabricated for the purpose of testing the method. The cylinders were made from 248 mm lengths of Plexiglas tubing. The inner and outer radii of the annular region between the cylinders were nominally 25.2 mm and 40.9 mm. The cylinders were from stock Plexiglas tubing, so the uncertainties in the radii were ± 1 mm. The outer cylinder was rotated with a rotation rate of 1.1 rad/s, with the inner cylinder stationary. In this example, employing a dye marker in glycerine, the fluid flow is nominally two-dimensional and the marker follows the flow. The image-flow velocity field, \mathbf{u} , as described in Sect. 1.1, can be accepted as a good representation of the fluid-flow velocity field, \mathbf{u}_f .

Images were recorded using a Texas Instruments Multi-camera MC-1134GN Multi-Mode B/W Camera. The data were stored digitally using an in-house multiple-frame grabber (12-bit A/D), designed by Dan Lang and Paul Dimotakis of GALCIT, set to record up to 28 of the 1134×468 pixel gray-level images from the camera, spaced by 100 msec (adjustable between 33 and 267 msec). Because the horizontal and vertical spacing of the pixels were not equal on this CCD, grid spacings and image correlation volumes with a ratio of 1:1.74 (vertical:horizontal pixels) were used to yield a uniform spacing of the data in the real image plane. Flow visualization was performed by randomly distributing red food coloring (dye) on the surface of the fluid. To provide backlighting for the dye marker, the fluid beneath the surface contained a translucent white suspension of 3 μm aluminum oxide (Al_2O_3) particles in glycerine. When illuminated from the side, this provided nearly uniform white backlighting for the dye being imaged on the surface. Because of the depth of field of the

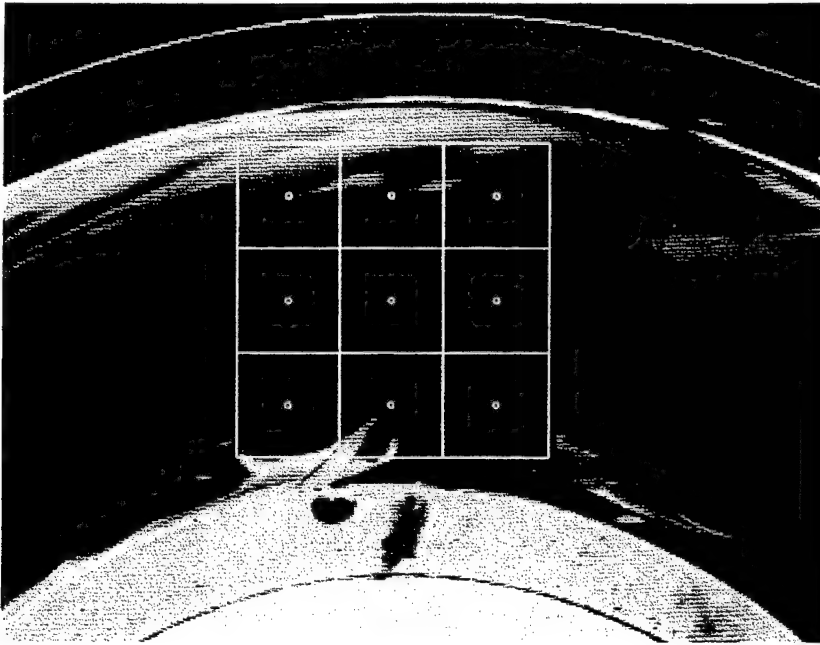


Fig. 5. Initial placement of series expansion neighborhoods. Each square denotes a control volume. The small circle at the center of each control volume denotes the center, or control point, of a series expansion.

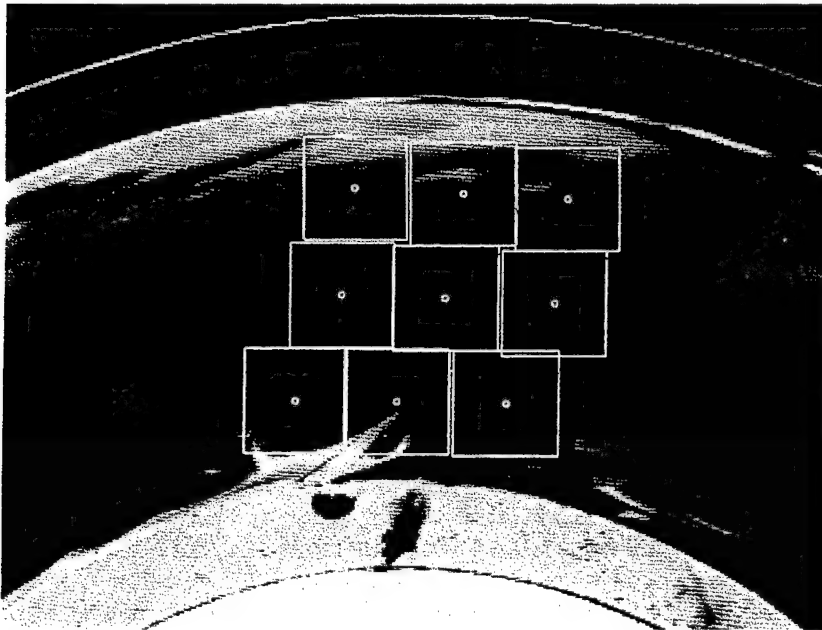


Fig. 6. Displacement of series neighborhoods, 100 ms later, allowing only translation.

imaging and the high density and uniform distribution of the aluminum oxide, scattering from individual particles was not detectable in the video images.

In the present investigations, only the outer cylinder was rotated, hence the velocity field can be written as (e.g., Schlichting 1979),

$$\frac{u_r(r, \theta)}{\Omega_0 r_0} = 0 \quad \text{and} \quad \frac{u_\theta(r, \theta)}{\Omega_0 r_0} = \frac{r/r_i - r_i/r}{r_0/r_i - r_i/r_0}, \quad (46)$$

where r , θ , u_r , and u_θ are the radial and angular positions and velocities, respectively, Ω_0 is the rotation rate of the outer cylinder, and r_i and r_0 are the inner and outer radii of the cylinders. In this flow, the divergence is zero, i.e.,

$$\nabla \cdot \mathbf{u} = 0, \quad (47)$$

and the vorticity is uniform, i.e.,

$$\omega_z(r, \theta) \equiv \mathbf{e}_z \cdot \nabla \times \mathbf{u}(r, \theta) = \frac{2\Omega_0}{1 - (r_i/r_0)^2}. \quad (48)$$

For this simple test case, nine correlation volumes spaced by 45 pixels vertically and 78.3 pixels horizontally, were used to capture the flow. The locations of the image correlation volumes at the initial time are depicted in Fig. 5. The results of the correlation process, allowing only displacement of the correlation volume (zeroth-order series expansion) is shown in Fig. 6. Figure 7 demonstrates how the flow is captured when the correlation process is extended to include higher-order terms in the series expansion. To quantify this improvement,

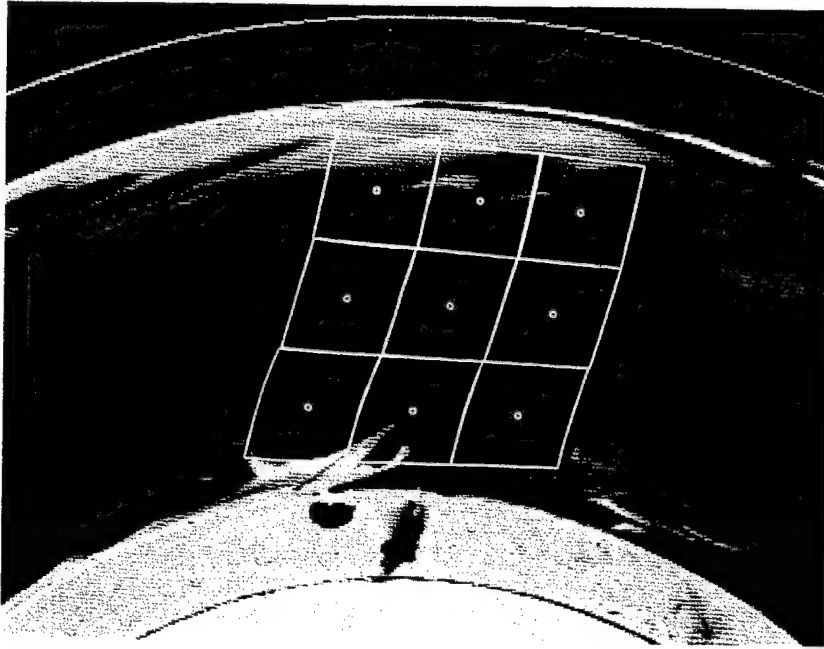


Fig. 7. Displacement of series neighborhoods estimated using higher order terms in the correlation process.

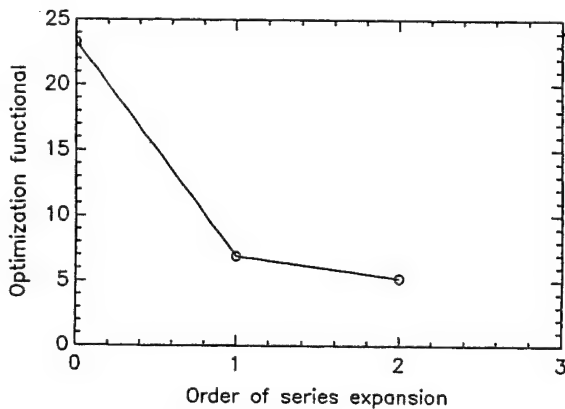


Fig. 8. Difference between Couette flow images under the mapping, quantified by the value of the optimization functional (arbitrary units), as a function of the number of terms in the two-dimensional series expansion.

the value of the minimization functional, not including the contribution of the constraints, is plotted in Fig. 8 as a function of the order of the series expansion. As can be seen, the greatest improvement is realized with the introduction of linear deformations in the correlation process.

The results using the correlation method are compared with the theoretical (analytical) two-dimensional values for the vorticity and divergence in Table 1. Since both the vorticity and the divergence are uniform in the analytical solution, only a single value is presented. The uncertainty in the theoretical vorticity is a result of the eccentricity of the cylinders used in the experiment. The uncertainties in the experimental values are one standard deviation. Because this flow has a nearly linear velocity profile, we saw only small changes in these estimates, beyond the linear order. Some of the "uncertainty" in the experimental results for the vorticity reflect the expected variations of the vorticity within the flow.

Table 1. Comparison of theoretical and experimental vorticity and divergence

	$\omega_z [\text{sec}^{-1}]$	$\nabla \cdot \mathbf{u} [\text{sec}^{-1}]$
Theoretical:	-3.6 ± 0.1	0.0
Experimental:	-3.6 ± 0.3	-0.1 ± 0.1

4

Cylinder-wake flow. Two-dimensional slices of three-dimensional data

In this section, we present the results of applying the two-dimensional correlations to two-dimensional slices of a three-dimensional flow in the wake of an impulsively started circular cylinder. Here, the cylinder is 1.75 cm in diameter and 45.5 cm long. It is drawn through a distribution of fluorescein dye in water at a speed of 1.27 cm/s. The Reynolds number in this case is,

$$Re \equiv \frac{Ud}{\nu} \approx 220, \quad (49)$$

where U is the cylinder speed, d is the cylinder diameter, and ν is the kinematic viscosity. The CCD camera and data acquisition system are the same as for the Couette flow test case (Sect. 3). Laser sheet illumination is provided by a Continuum model YG661-10 frequency-doubled YAG laser. The laser was operated at 532 nm, 300 mJ, 5 ns pulses width, at a rate of 10 Hz. The flow here is three-dimensional in both the velocity field and scalar distribution.

Figure 9–14 demonstrate the method on images of a vortical structure forming in the wake of the cylinder. These images were taken after the cylinder had travelled about 8 diameters. The image at the initial time is shown in Fig. 9, and 100 ms later, in Fig. 10. In this case, the series approximation used in

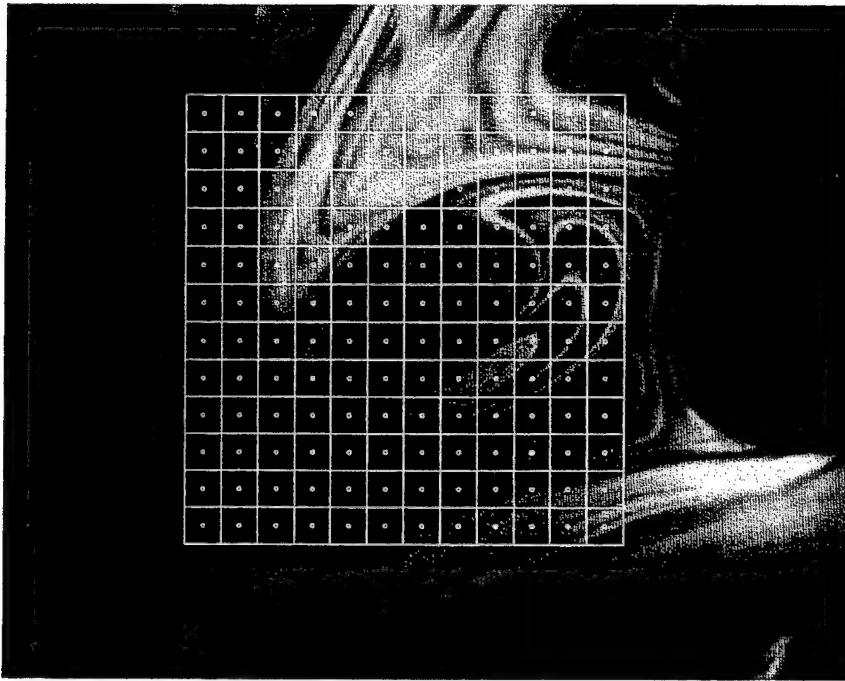


Fig. 9. Initial placement of series expansion neighborhoods. Each square denotes a control volume. The small circle at the center of each control volume denotes the center, or control point, of a series expansion.

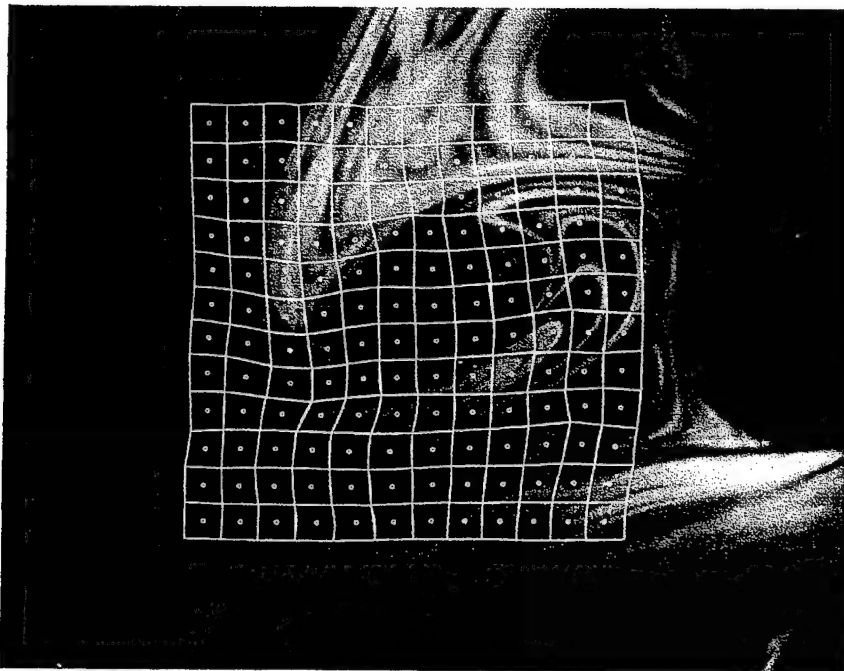


Fig. 10. Displacement of grid after 100 ms, estimated using the nonlinear correlation process.

the correlation process was expanded to quadratic order. Figure 11 shows the displacement of the centers of the correlation volumes.

The two-dimensional vorticity is displayed in Fig. 12. A large vortical region can be seen in the wake of the cylinder. The two-dimensional divergence, presented in Fig. 13, exposes the three-dimensionality in the flow. Figure 14 plots the second invariant of the rate-of-strain tensor. Note the region of large strain (rate) that seems to follow the periphery of the large vortical structure. This could be a region of vorticity, from the previously shed vortical structure, that is being strained around the current one.

As a general observation, an important issue arises when imaging a two-dimensional (planar) slice of a three-dimensional field of a continuous scalar, $c(\mathbf{x}, t)$, as in the previous example. An out-of-plane component of the fluid-flow velocity, \mathbf{u}_f , coupled with an out-of-plane component of the scalar gradient, $\nabla c(\mathbf{x}, t)$, will contribute to the in-plane image-flow velocity \mathbf{u} . In this case, the equation for the in-plane image flow can still be written as,

$$\frac{\partial c}{\partial t} + \mathbf{u} \cdot \frac{\partial c}{\partial \mathbf{x}} \equiv \frac{\partial c}{\partial t} + u \frac{\partial c}{\partial x} + v \frac{\partial c}{\partial y} = 0, \quad (50)$$

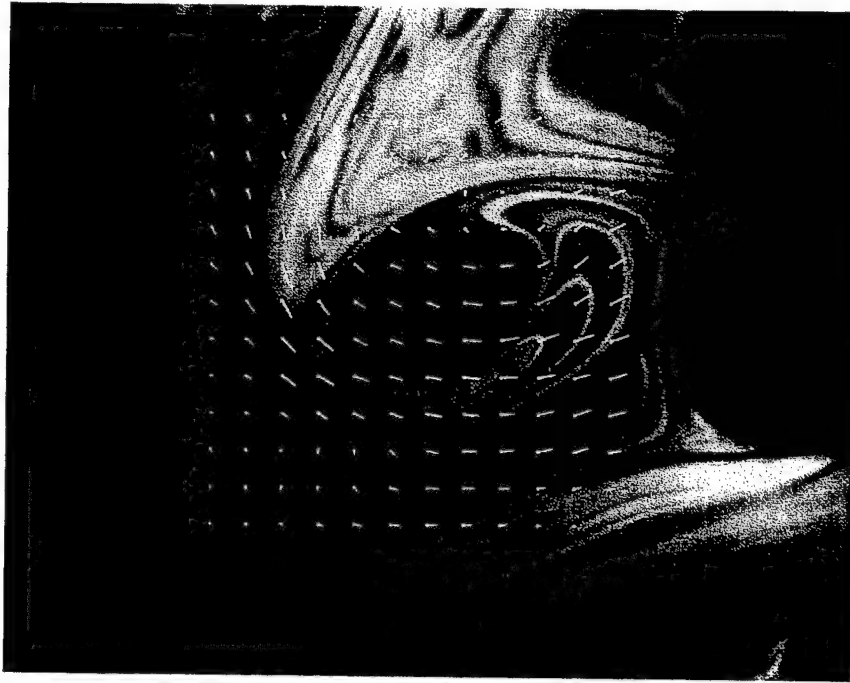


Fig. 11. Displacement of centers of grid over 100 ms.

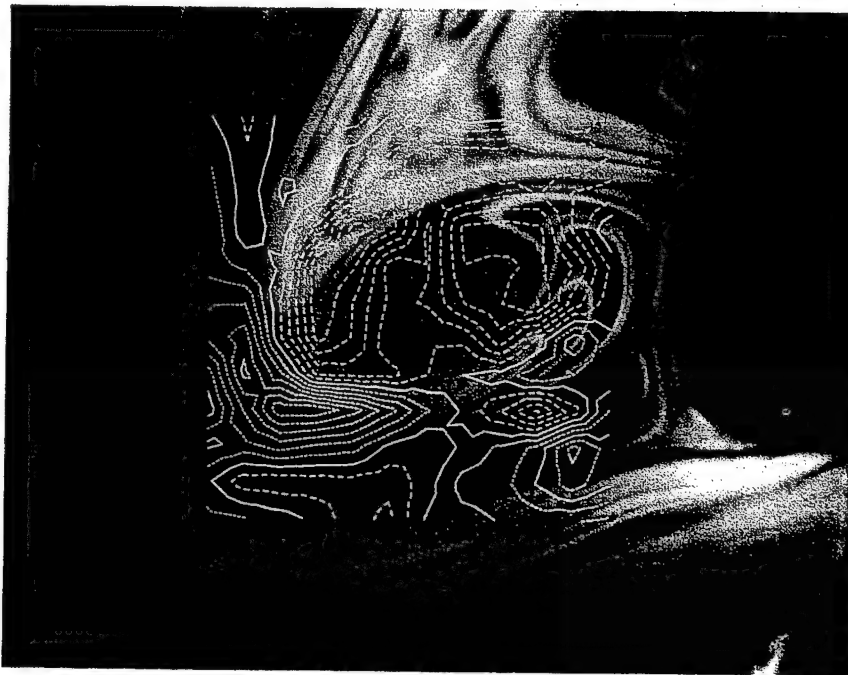


Fig. 12. Contours of constant plane-normal vorticity, $\nabla \times \mathbf{u} = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$. Solid contours denote zero-vorticity. Long dashes denote positive values and short dashes negative. Contours spaced by 0.5 s^{-1} .

where we have assumed that the image irradiance $E(\mathbf{x}, t)$ is proportional to the two-dimensional slice of the scalar concentration, $c(\mathbf{x}, t)$, and where the in-plane image-flow velocity, $\mathbf{u} = (u, v) = d\xi/dt$ (Eq. (11)), is the one derived from the minimization function, as described above.

Considering the transport of the three-dimensional iso-scalar surfaces (see Fig. 15), we find that the two-dimensional $\mathbf{u} = (u, v)$ in-plane image-flow velocity components are related to the three-dimensional

$\mathbf{u}_t = (u_t, v_t, w_t)$ fluid-flow velocity and the three-dimensional scalar gradient components. In particular, we have,

$$u = u_t + w_t \frac{\partial c}{\partial z} \frac{\partial c / \partial x}{(\partial c / \partial x)^2 + (\partial c / \partial y)^2},$$

$$v = v_t + w_t \frac{\partial c}{\partial z} \frac{\partial c / \partial y}{(\partial c / \partial x)^2 + (\partial c / \partial y)^2}.$$
(51)

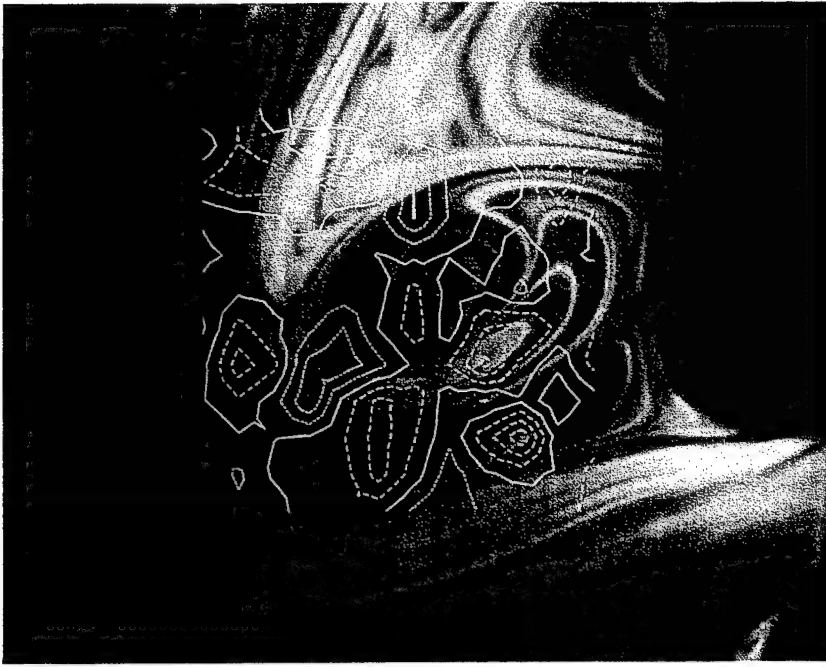


Fig. 13. Contours of constant divergence,
 $\nabla \cdot \mathbf{u} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}$. As in Fig. 12, the contours are
 spaced by 0.5 s^{-1} .

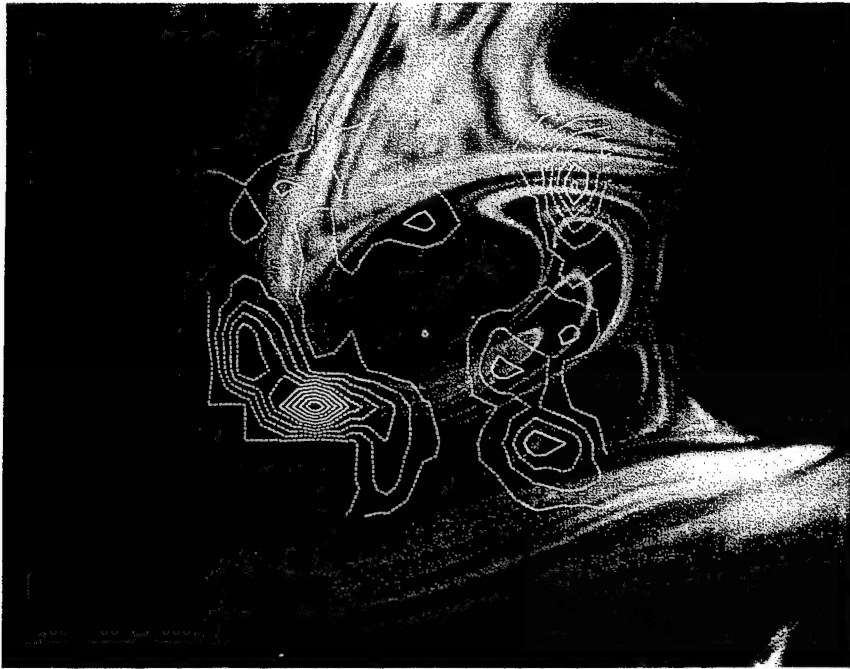


Fig. 14. Contours of constant second
 invariant of the rate-of-strain tensor (Eq. 45),

$$q_s = -\frac{1}{2} \left[\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial v}{\partial y} \right)^2 + \frac{1}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right)^2 \right]$$

Contours spaced by 0.5 s^{-2} .

As can be seen, by substituting Eq. (51) in Eq. (50), these relations recover the three-dimensional transport equation for a conserved scalar field $c(\mathbf{x}, t)$, in the case of negligible diffusion, i.e.,

$$\frac{\partial c}{\partial t} + \mathbf{u}_f \cdot \frac{\partial c}{\partial \mathbf{x}} \equiv \frac{\partial c}{\partial t} + u_f \frac{\partial c}{\partial x} + v_f \frac{\partial c}{\partial y} + w_f \frac{\partial c}{\partial z} = 0.$$

These results provide us with the criteria for when the in-plane image-flow velocity can be regarded as a good

approximation to the in-plane fluid-flow velocity. In particular, we will have,

$$u \simeq u_f, \quad \text{if} \quad \frac{u_f}{w_f} \gg \left(\frac{\partial c / \partial z}{(\partial c / \partial x)^2 + (\partial c / \partial y)^2} \right) \frac{\partial c}{\partial x}, \quad (52a)$$

and

$$v \simeq v_f, \quad \text{if} \quad \frac{v_f}{w_f} \gg \left(\frac{\partial c / \partial z}{(\partial c / \partial x)^2 + (\partial c / \partial y)^2} \right) \frac{\partial c}{\partial y}, \quad (52b)$$

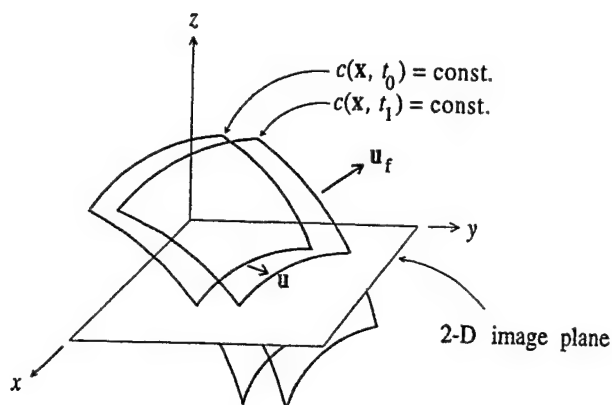


Fig. 15. Two-dimensional slices of a three-dimensional scalar field $c(x, t)$. u_f indicates the 3-D fluid-flow velocity and u is the resulting two-dimensional in-plane image-flow velocity.

We can see that if w_f is small, or if $\partial c / \partial z$ is small, or both, by the measure in Eq. (52), then the in-plane image-flow velocity field can be accepted as a good representation of the in-plane fluid-flow velocity field.

Finally, since this method estimates the in-plane image-flow velocity, u , and not the fluid-flow velocity u_f , its application to two-dimensional image slices of three-dimensional scalar field data is identical to its application to two-dimensional scalar data.

5

Voyager 2 images of Jupiter

The method is also illustrated on a pair of the images of the atmospheric dynamics of Jupiter taken by Voyager 2. These images were taken from the "Voyager Time-Lapse, Cylindrical-Projection Jupiter Mosaics," by Avis & Collins (1983). 640×350 pixel subimages of rotations 349 and 350 were used in the correlation process. The subimage spans 168° to 97° longitude and 0° to -46° latitude (the equator is at the top of the image). The subimage from rotation 349 is shown in Fig. 16, with an overlay of the initial placement of the correlation volume neighborhoods. The vertical line on the left is a reference line which is to be deformed using the mean zonal velocities of Jupiter from Limaye (1986). Figure 17 shows the same region, one rotation later, with the associated grid deformed by the nonlinear correlation method. On the left is the reference line from Fig. 16, carried by the mean zonal flow. The displacement of the centers of the correlation volumes is shown in Fig. 18. As discussed Sect. 1.1, the motion of the imaged quantity in the Jovian atmosphere does not follow any simple equation of motion and the derived image-flow velocity field can be expected to be a relatively poor representation of the fluid-flow velocity field.

6

Conclusions

Series expansions of image displacement, in conjunction with a global nonlinear correlation method, can be used to estimate

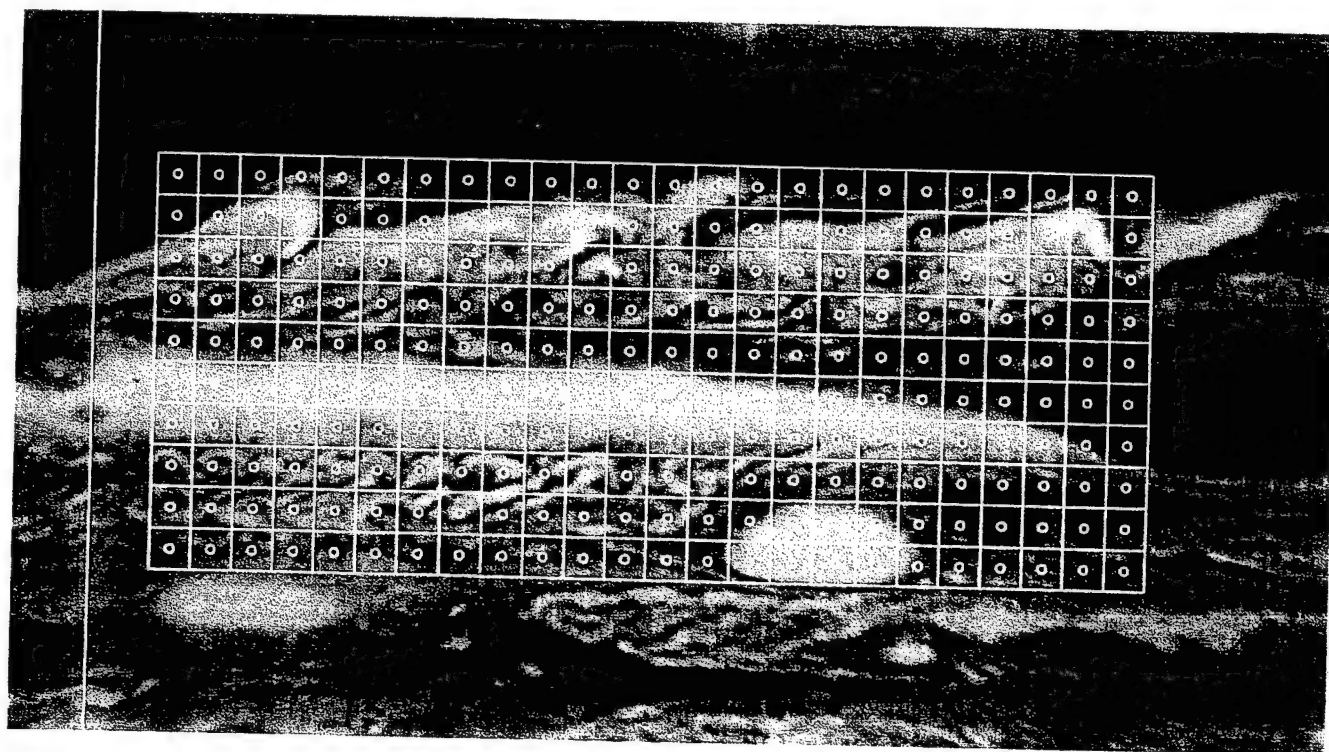


Fig. 16. Initial placement of correlation volumes overlaid on a sub-image of rotation 349 from the "Voyager Time-Lapse, Cylindrical-Projection Jupiter Mosaics." Each square of the grid denotes a correlation volume. The vertical line on the left is a reference line to be carried with the mean zonal velocity of Jupiter (see Fig. 17).

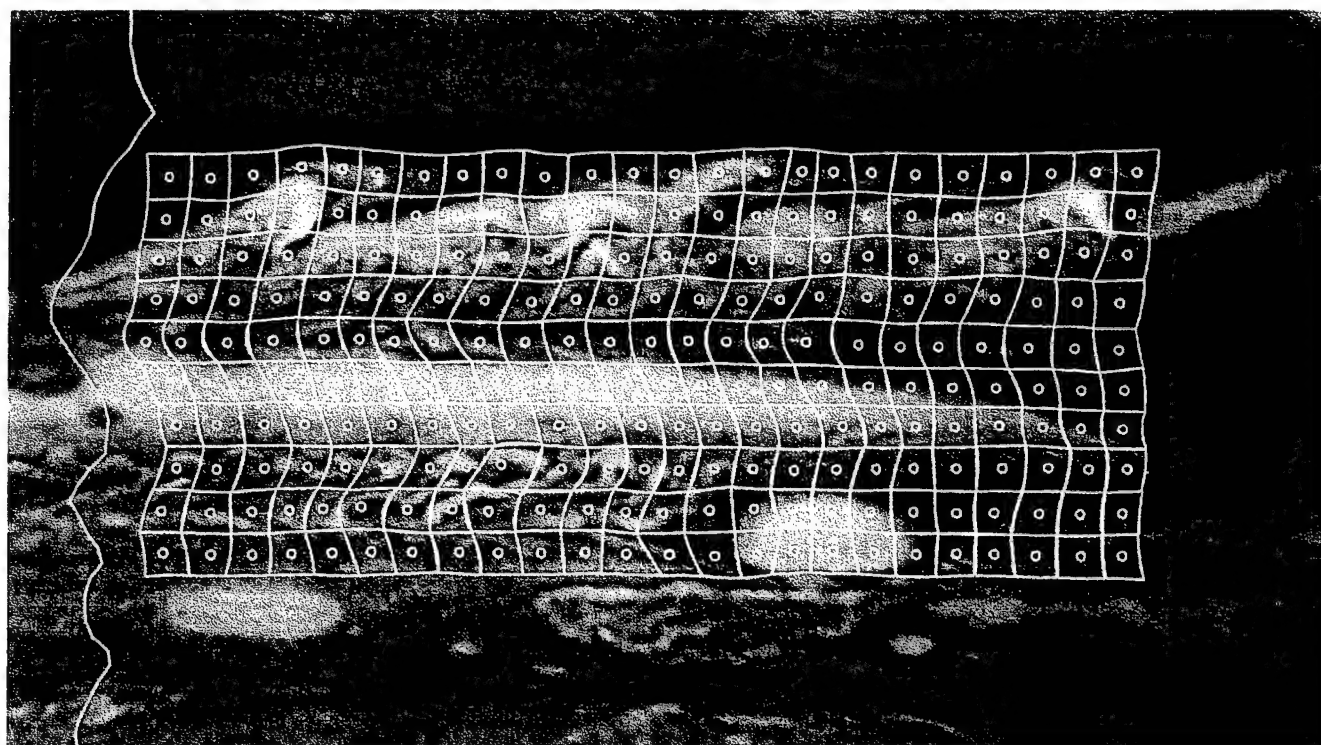


Fig. 17. Deformation of the correlation volumes (see Fig. 16), after one rotation. The line on the left was deformed from the vertical line in Fig. 16 using the mean zonal velocity of Jupiter.

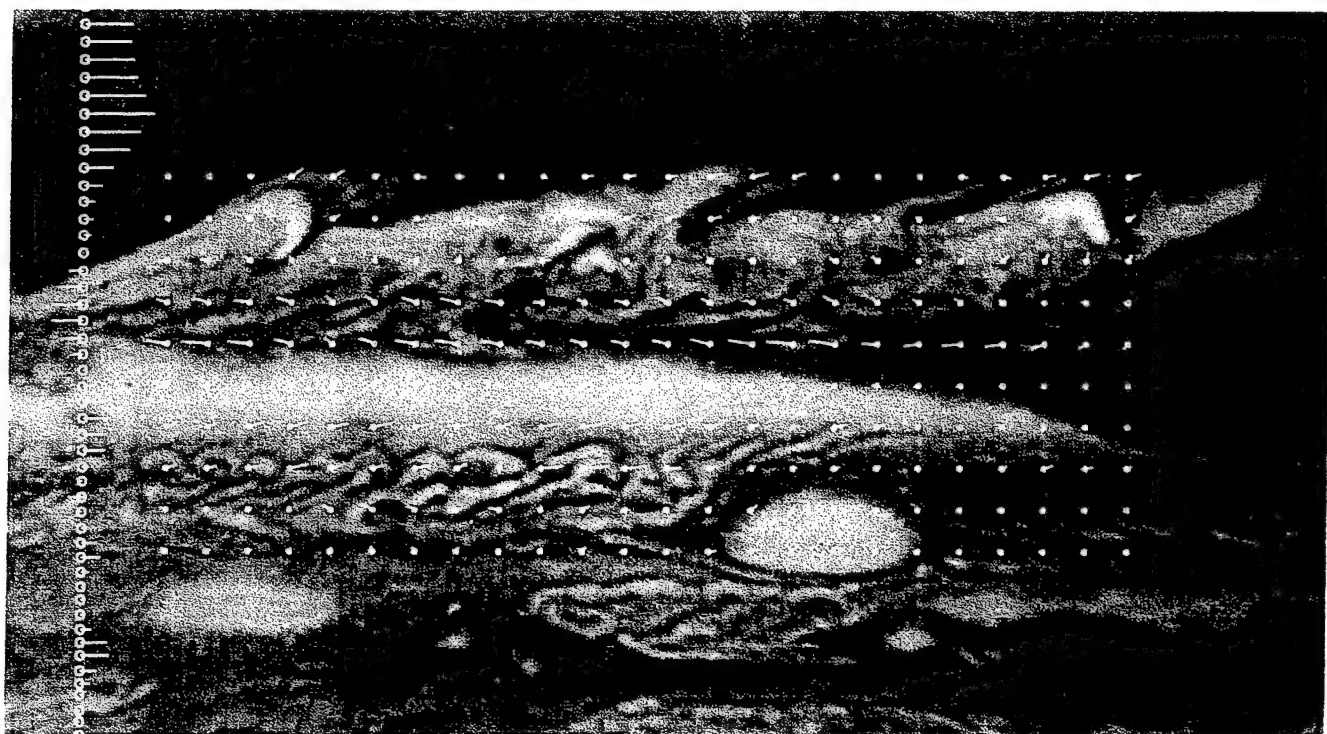


Fig. 18. Displacement of grid control points (centers of correlation volumes) after one rotation. The lines on the left denote the displacement via the mean zonal flow of Jupiter.

fluid velocities, and velocity gradients, from pairs of convected scalar images. It is shown that increasing the order of the expansion can improve the accuracy of the results. The proposed method does not require, or preclude, discrete particles and may also be used in situations where there is a natural marker already in the flow, e.g., successive images of species concentration can be used to estimate velocities in compressible flows. The method is developed for three-dimensional data sets and demonstrated on two-dimensional images of fluid flow.

References

- Adrian RJ (1991) Particle-imaging techniques for experimental fluid mechanics. *Ann Rev Fluid Mech* 23: 261–304
- Avis CC; Collins SA (1983) Voyager time-lapse, cylindrical-projection mosaics. Report D-541, Jet Propulsion Laboratory, Pasadena, CA
- Bindschadler RA; Scambos TA (1991) Satellite-Image-Derived Velocity field of an Antarctic Ice Stream. *Science* 252: 181–344
- Brand AJ; Hesselink L (1990) Auto-correlation measurements in three-dimensional fluid flow datasets. *Exp Fluids* 10: 55–59
- Cantwell BJ (1992) Exact solution of a restricted Euler equation for the velocity gradient tensor. *Phys Fluids A* 4 (4): 782–793
- Dahm WJA; Southerland KB; Buch KA (1991) Direct, high resolution, four-dimensional measurements of the fine scale structure of $Sc \gg 1$ molecular mixing in turbulent flows. *Phys Fluids A* 3(5, Pt. 2): 1115–1127
- Dahm WJA; Su LK; Southerland KB (1992) A scalar imaging velocimetry technique for fully resolved four-dimensional vector velocity field measurements in turbulent flows. *Phys Fluids A* 4(10): 2191–2206
- Falco RE; Chu CC (1987) Measurement of two-dimensional fluid dynamic quantities using photochromic grid tracing technique. *SPIE* 814: 706–710
- Gibson JJ (1950) *The Preception of the Visual World*. Houghton Mifflin, Boston
- Hildreth EC (1984) *The Measurement of Visual Motion*. The MIT Press
- Hildreth EC; Koch C (1987) The Analysis of visual-motion – from Computational theory to neuronal mechanisms. *Ann Rev Neur* 10(): 477–533
- Horn BKP; Schunck BG (1981) Determining optical flow. *Artificial Intelligence* 17: 185–203
- Huang HT; Fiedler HE; Wang JJ (1993a) Limitation and improvement of PIV; Part I: Limitation of conventional techniques due to deformation of particle image patterns. *Exp Fluids* 15: 168–174
- Huang HT; Fiedler HE; Wang JJ (1993b) Limitation and improvement of PIV; Part II: Particle image distortion, a novel technique. *Exp Fluids* 15: 263–273
- Hutchinson J; Koch C; Luo J; Mead C (1988) Computing motion using analog and binary resistive networks. *Computer* 21(3): 52–63
- Limaye SS (1986) Jupiter: New estimates of the mean zonal flow at the cloud level. *Icarus* 65: 335–352
- McNeil SR; Peters WH; Sutton MA (1987) Estimation of stress intensity factor by digital image correlation. *Engineering Fracture Mechanics* 28(1): 101–112
- Miles RB; Connors JJ; Markovitz EC; Howard PJ; Roth GJ (1989) Instantaneous profiles and turbulence statistics of supersonic free shear layers by Raman excitation plus laser-induced electronic fluorescence (Relief) velocity tagging of oxygen. *Exp Fluids* 8: 17–24
- Murray DW; Buxton BF (1990) *Experiments in the machine interpretation of visual motion*. The MIT Press
- Pearlstein AJ; Carpenter BN (1993) On the determination of solenoidal or compressible velocity fields from measurements of passive or reactive scalars. *Phys. Fluids* (accepted)
- Press WH; Flannery BP; Teukolsky AA; Vetterling WT (1992) *Numerical Recipes. The Art of Scientific Computing*. Second Edition. Cambridge Univ. Press
- Schlichting H (Tr J Kestin) (1979) *Boundary-layer theory*. Seventh Edition, McGraw-Hill Pub Co
- Terzopoulos D; Metaxas D (1991) Dynamic 3D models with local and global deformations: deformable superquadrics. *IEEE Trans on Patt Anal and Mach Int* 13(7): 703–714
- Wallach H (1976) On perceived identity: I. The direction and motion of straight lines. In: *On Perception*, ed. by H. Wallach, Quadrangle, New York
- Willert CE; Gharib M (1991) Digital particle image velocimetry. *Exp Fluids* 10: 181–193

Pre-Publication: To be presented at:
IS&T/SPIE Symposium on Electronic Imaging: Science & Technology, 2/4-10/95, San Jose, CA. Document #2415-01

Sandbox CCDs

Jim Janesick, Tom Elliott
Jet Propulsion Laboratory
California Institute of California
4800 Oak Grove, Pasadena, CA. 91109

Rusty Winzenread

EG&G Reticon
345 Potrero Ave., Sunnyvale, CA. 94086

Jeff Pinter

Loral Fairchild Imaging Sensors
14251A Chambers Rd., Tustin, CA. 92680

Rudolph Dyck

Loral Fairchild Imaging Sensors
1801 McCarthy Blvd., Milpitas, CA. 95035

ABSTRACT

Seven new CCDs are presented. The devices will be used in a variety of applications ranging from generating color cinema movies to adaptive optics camera systems to compensate for atmospheric turbulence at major astronomical observatories. This paper highlights areas of design, fabrication, and operation techniques to achieve state-of-the-art performance. We discuss current limitations of CCD technology for several key parameters.

Keywords: Charge-coupled devices, solid state imagers, charge transfer efficiency, dynamic range, quantum efficiency, dark current, well capacity, read noise, x-ray, ultra-violet, high speed readout.

1. INTRODUCTION

CCD technology, now celebrating its twenty-fifth birthday, has matured to a remarkable degree. Through much of the 1970s, CCD fabrication was beset by countless difficulties, including contamination, variable and incompletely understood processing steps, rudimentary design tools, deficient modeling of device performance and absolute characterization testing techniques. As a consequence, a great deal of trial and error was involved. Technological progress took place despite these instabilities and unknowns but the price was high. As an example, two early JPL CCD programs, the Galileo Orbiter Solid State Imager and the initial Hubble Space Telescope Wide Field/Planetary Camera (WF/PC) entailed a total of approximately 150 production lots to develop and produce suitable CCDs.

In contrast, today it is not uncommon for a new custom CCD to be successfully produced on the first attempt. This maturation has enabled the production of CCDs at prices that are more widely affordable and has spawned an explosion in CCD applications. Some new applications include HDTV cameras, consumer electronics (e.g., camcorders), biological x-ray microscopes, medical x-ray cameras (e.g., spot mammography), dental x-ray cameras, document and graphic archiving scanners, ultra-large format astronomical imaging and spectroscopy cameras, modest-format amateur astronomy cameras, space-borne imaging camera systems, 35 mm still and motion picture cameras, industrial robotic cameras and PC cameras and scanners.

Indicators of this maturity are numerous:

- 1). Charge transfer efficiency ((0.9999995 for a 1620 e⁻ signal) and production yield have progressed to the point where CCDs having 10⁸ pixels are being planned. Device shorts have been reduced where it is now feasible to build wafer-scale arrays. The largest CCD (in pixel count) that could be theoretically fabricated today is a 16k x 21k, 5-micron pixel CCD (fabricated on a six-inch wafer).

- 2). Readout noise is as low as 2-3 e^- (rms) for conventional on-chip amplifiers. Values well below one e^- are achieved with floating gate amplifiers. Well capacity has been improved almost a factor of ten compared to CCDs fabricated 15 years ago. Dynamic range for some CCDs today is $> 10^6$.
- 3). The quantum efficiency (QE) of these devices is remarkably high (> 0.5) over unprecedented range of wavelengths, approximately 1-10,000 Å. One can now routinely obtain thinned, back illuminated devices with QEs in excess of 70-90 % from 380 to > 700 nm. New anti-reflection (AR) coatings are being developed to extend the useful response of the CCD well into the UV. One such coating will yield 50 % at 300 nm and > 20 % at 200 nm. Phosphor coatings applied directly to the CCD extend sensitivity into the hard x-ray regime (0.1-1 Å) further exceeding its useful range.

In view of this maturity, it is perhaps surprising to note the many areas in which CCD technology opportunities remain. For example, it appears that the potential for high speed CCD ($> 10^9$ pixels/sec) remains largely untapped. This delay in large part can be attributed to the lack of availability of high speed computers with lots of memory and relatively easy-to-use image processing software packages. Today CCD groups are designing 1024 x 1024 sensors that readout at 1000 frames/sec (over a billion pixels/sec). Similarly, the maximization of dynamic range for small pixels (< 10 microns) appear to be far from complete.

The authors represent three organizations that have participated in aggressively pushing the limits of CCD technology. Through several recent CCD development programs, we have developed a set of design and production techniques that reliably produces CCDs of high scientific performance. These programs include: the Cassini cameras that are to be launched towards Saturn in 1998, the replacement Hubble Space Telescope that was installed on-orbit in December 1993, the Multi-angle Imaging Spectro Radiometer (MISR) that is to be launched as part of NASA's Earth Observing System, an astronomical x-ray camera that is to study the distant x-ray background radiation, a high speed 1024 x 512 camera that records fluid dynamical phenomena by recording two images separated in time by only 200 ns with read noise $< 5 e^-$, and a high speed/low noise CCD used in adaptive optics systems. All of these CCDs share common design and fabrication characteristics. Collectively, these characteristics represent an excellent baseline from which to investigate additional technology problems and develop additional custom CCDs.

The "Sandbox lot" is an effort to capitalize upon this technology maturity to reduce the cost of CCD development and production. In doing so, our approach enables the development of custom CCDs whose cost would formerly have been unaffordable and also research of important CCD issues (such as dynamic range optimization) at greatly reduced cost. These savings are realized by combining several devices on a single development lot, thus dividing the production costs among several customers.

This paper consists of two major sections. The first section describes seven specific CCDs that are being developed in this manner. The second section discusses the performance, design and production issues that are involved in developing these seven devices. Many of the designs incorporate features that further push the limits of CCD technology. Key areas that are discussed include: dark current generation, charge transfer, well capacity, high speed readout, QE, read noise, and high energy radiation tolerance.

The first six Sandbox CCDs listed above will be fabricated in one single lot run at Reticon. The seventh CCD (Big CIT) will occupy an entire 4-inch wafer and will be fabricated at Loral Fairchild. The wafer runs fabricated at both Reticon and Loral Fairchild will be made with a standard three poly, two metal NMOS CCD process. Two of the sensors (Advanced Camera and Big CIT) are rear-illuminated and therefore will require thinning and backside coatings (AR and phosphor). These tasks will be performed at Reticon (Ref. 2). Wafer testing, backside accumulation, screening and calibration tests will be done at JPL. All Sandbox CCD designs were created at JPL in collaboration with Reticon and Loral Fairchild design engineers.

2. SANDBOX CCD DESCRIPTIONS

2.1 Cinema CCD

The Cinema CCD is a first generation 4096 x 4096, 3-phase, 9-micron pixel, frame transfer, color CCD. The sensor is intended to replace photographic film used in Hollywood moving picture cameras. The new CCD offers several new design challenges.

For example, the CCD will be readout at 30 frames/sec at an effective pixel rate of a quarter billion pixels per second. To achieve such rates the CCD is divided up into 32 sections for parallel readout. The array itself is split into four 1024 (V) x 4096 (H) regions forming two image and two storage sections. The split image is transferred towards the top and bottom of the array. From each storage region there are 16 horizontal registers that report to sixteen 3-stage on-chip amplifiers (Figure 1a). Each channel reads at approximately 8 M pixels/sec. Each pixel will be digitized in the camera 12-bits deep. A three-chip approach will be employed to obtain color information.

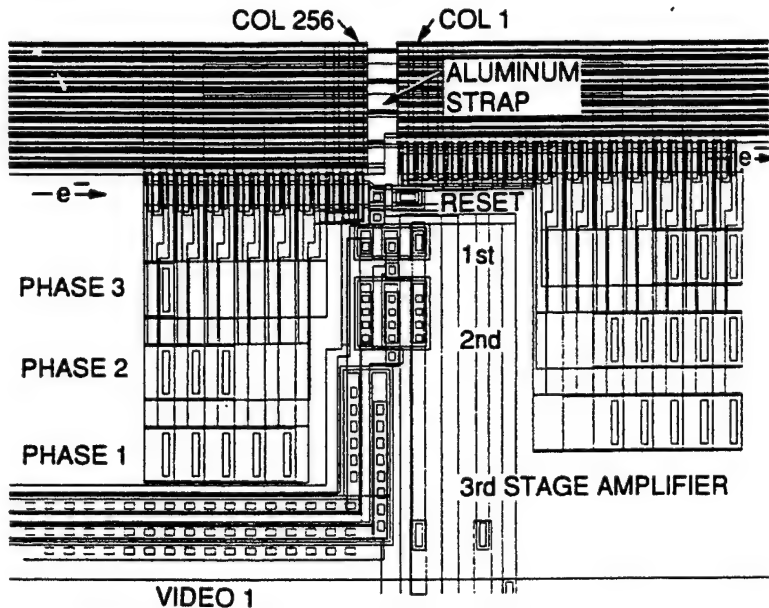


Figure 1a. The Cinema CCD three-stage output amplifier. Thirty-two amplifiers are used to read the sensor at approximately 8 Mpixels / sec / channel. The illustration shows staggered horizontal registers to access sense node. An aluminum strap is bussed into the array to improve high speed drive characteristics. Image processing performed by a large computer will remove his artifact from images generated.

Each image generated by the Cinema CCD will be processed by a large computer system. Some image processing requirements performed are: (1) gain/offset correction for each channel, (2) cosmetic removal, (3) image compression, and (4) color correction and alignment between the three CCDs. Since images are in digital format, "special effects" can be added to processed images at the same time.

2.2 Advanced Camera CCD

The Advanced Camera CCD is a high performance 4096 (V) x 4096 (H), 3-phase, 9-micron pixel, full frame, backside illuminated, CCD. The sensor is intended to be used on a third generation Hubble Space Telescope instrument. The array and two horizontal registers are split allowing 4 channel parallel readout. Two floating diffusion and two floating gate (Skipper) amplifiers are employed for ultra-low-noise performance (sub-electron rms). The device will be thinned and use AR and phosphor coatings to cover a wide wavelength range (1100 - 11000 Å). The backside will be accumulated with a molecular beam epitaxial process. For the Sandbox lot, this CCD is being built with a format of 4098 x 3071.

2.3 MACH II CCD

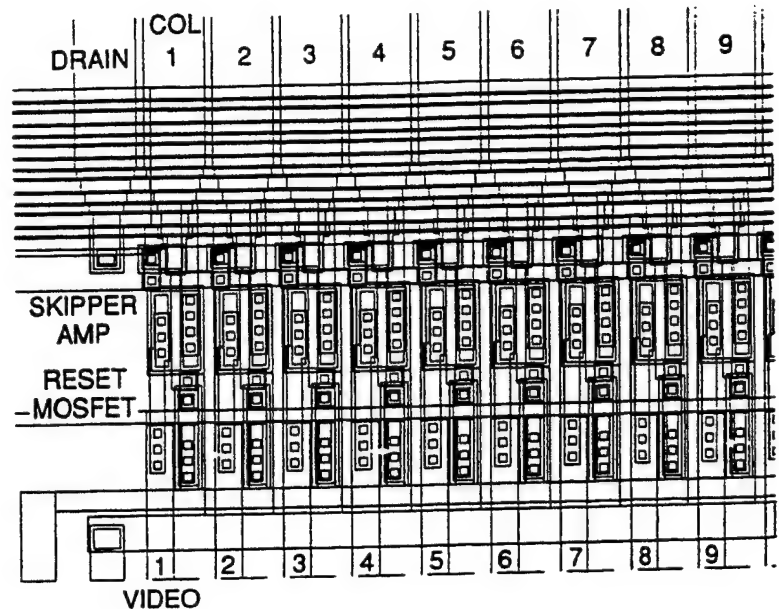
The MACH II CCD is a second generation 1024 x 1024, 6-phase, 36(V) x 18(H)-micron pixel, interlaced-frame transfer CCD. The sensor will be used to take two sequential 512 (V) x 1024 (H) images separated in time by less than 1-micro-sec. Phases 4, 5, and 6 are masked off with a light shield serving as a storage region. The first image is collected in phases 1-3 and then quickly transferred (<1-micro-sec) into storage phases 4-6. A second image is then taken. The two interlaced images are then readout slow-scan (50 kpixels/sec) to achieve 4 e- noise performance. Six-phase design is employed for high QE performance, low optical cross-talk between the two images and high speed operation. The CCD also incorporates a high speed readout channel for fast-scan applications where low-noise performance is not required.

2.4 Adapt II CCD

The Adapt II CCD is a second generation 128(V) x 64 (H), 6-phase, 36-micron pixel, frame store, CCD. The CCD will be used

for wavefront sensing in Adaptive Optics camera systems. The device will scan at 2000 frames/sec at a read noise < 3 e- rms. Sixty-four amplifiers are provided, one amplifier per column (Figure 1b). The amplifiers are Skipper type to allow for multiple sampling and low-noise.

Figure 1b. Adapt II CCD output Skipper amplifiers. Sixty-four amplifiers are provided, one for each column. Column pitch is 36-microns.. Six-phase (6-microns / phase) clocking is employed for high speed frame transfer operation.



2.5 Pluto Flyby CCD

The Pluto Flyby CCD is a first generation 2048 (V) x 1024 (H), 3-phase, 9-micron pixel, frame store, CCD. The CCD will potentially be used in a NASA planetary mission to planet Pluto. One simple floating diffusion amplifier is used to readout the array.

2.6 Circulator CCD

The Circulator CCD is a 16 pixel, circular, test CCD (Figure 1c). The device is intended to interrogate a single electron. To do this, a Skipper amplifier takes a sample each time the electron makes one revolution around the circuit. The amplifier and signal channel have been designed for the lowest noise possible (e.g., signal channel width is only 5-microns wide to reduce readout capacitance). Thousands of samples can be taken to reduce the noise below 0.1 e- rms. At the center of the array is a diode which can be slightly forward biased to emit photons and generate charge in the signal channel if desired. We intend to inject a single electron thermally by controlling operating temperature. A dump gate and drain are also provided to erase charge in the channel before the experiment is performed.

2.7 Big CIT CCD

The Big CIT (California Institute of Technology) CCD is a 4096 x 4096, 3-phase, 15-micron, full frame, backside illuminated, CCD. The device will be used in ground based astronomical applications. The CCD is configured similar to the Advanced Camera CCD (i.e., quad array), allowing four channel readout if desired. The CCD will be one of the largest CCDs manufactured today.

3. PERFORMANCE CONSIDERATIONS

3.1 Dark current

A critical CCD parameter is thermally generated dark current. For CCD imagers there are three main sources of dark current.

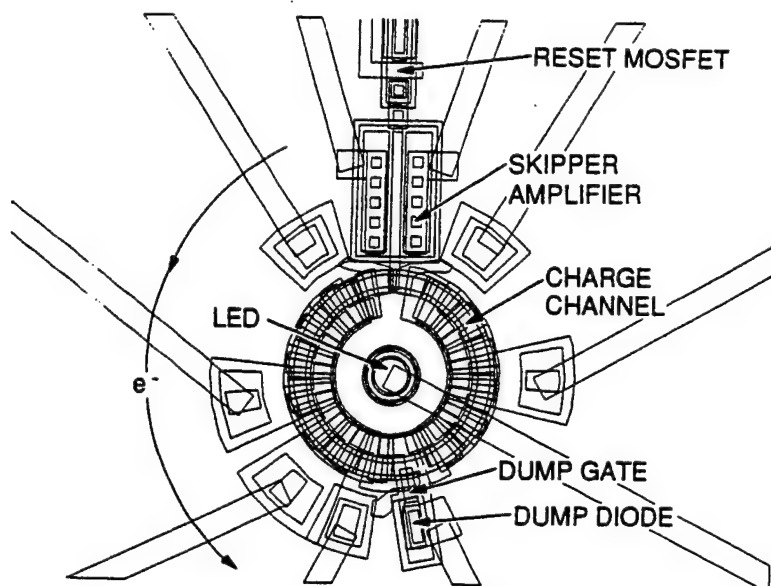


Figure 1c. Circulator CCD showing 16 three-phase pixels. Charge is readout nondestructively by a Skipper amplifier to achieve sub electron noise performance. A dump gate and dump diode are used to quickly erase charge in signal channel. A LED is provided to perform photon transfer calibration. The purpose of the device is to detect and locate a single thermal electron when generated.

These are: (1) thermal generation and diffusion in the neutral bulk, (2) thermal generation in the depletion region and (3) thermal generation due to surface states at the Si-SiO₂ interface. Of these sources, the contribution from surface states has been the dominant contributor for CCDs. Surface dark current varies significantly among manufacturers depending on processing details associated with oxide growth and surface passivation. Recent dark current tests show that dark levels vary widely for CCD manufacturers from 60 pA/cm² to 10 nA/cm² for noninverted CCDs (unless otherwise specified dark current figures in this paper are referenced to room temperature operation). Still higher generation rates (10,000 nA/cm²) are exhibited for backside illuminated CCDs when not properly accumulated.

Dark current generation at the Si-SiO₂ interface is primarily determined by two factors, namely the density of interface states and the density of free carriers (holes or electrons) that populate the interface. Electrons can thermally "hop" from the valence band to an interface state and into the conduction band as a free electron. The electron is then collected in a potential well as real signal. The presence of free carriers can fill these states inhibiting the hopping conduction process and in turn substantially reduce the dark generation rate. Noninverted CCDs maximize the amount of dark current generated since the interface is completely depleted of free carriers. However, when the CCD is inverted (e.g., MPP), holes from the channel stops collect and populate the interface eliminating surface dark current. The only remaining dark current is produced in the bulk silicon.

Bulk dark current (i.e., MPP operation) also varies significantly from manufacturer to manufacturer. Levels as low as 2 pA/cm² and as high as 1 nA/cm² have been measured. Dark rate may also vary significantly from lot to lot for a single manufacturer using a single process. For example, three consecutive lots were fabricated for the recent WF/PC II project. Dark rates for these CCDs ranged from 20 - 400 pA/cm². The difference noted was related to the quality of the silicon wafers used and bulk state density.

Bulk states are lattice imperfections or impurities of some kind (e.g., gold). High quality silicon implies that bulk state density is very low. Bulk states generate dark current also via "hopping conduction" by jumping through states in the band-gap. When a CCD is "bulk state limited" individual spikes often dominate the dark floor seen. For example, Figure 2a shows a dark current line stacking plot generated by a 1024 x 1024 Loral Fairchild MPP CCD. The average dark floor measured for the sensor is < 10 pA/cm². Note that each spike generates approximately the same amount of charge indicating a specific impurity is at work. Some pixels exhibit multiple spikes explaining why several distinct signal levels are seen in the plot. The dark current pedestal on which the spikes ride is believed to be in part generated in the "birds beak" near the channel stop region where inversion is not fully achieved (refer to Section 7). Also, surface dark current is generated each time a line is transferred since the phases must come out of inversion for a brief period of time.

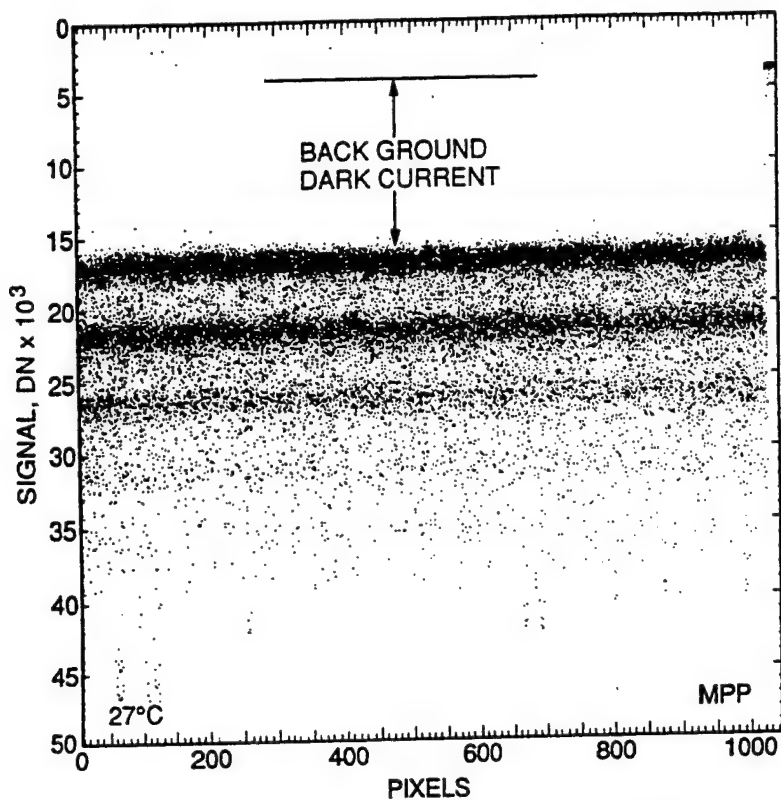


Figure 2a. Dark spike pattern generated by a MPP CCD. Spikes are associated with a specific bulk state that generates the same amount of charge per unit time. Some pixels contain multiple spikes resulting in discrete signal levels shown. Background dark current is generated from several sources including surface dark current when phases come out of inversion during line transfer.

The CCD tested in Figure 2a can integrate charge for six minutes at room temperature before saturating. Figure 2b plots average dark current build up as a function of lines readout (each line takes 23.5 ms). Partial inverted and MPP responses are shown. After one frame readout (23 seconds) the MPP dark floor has only risen by 3000 e^- .

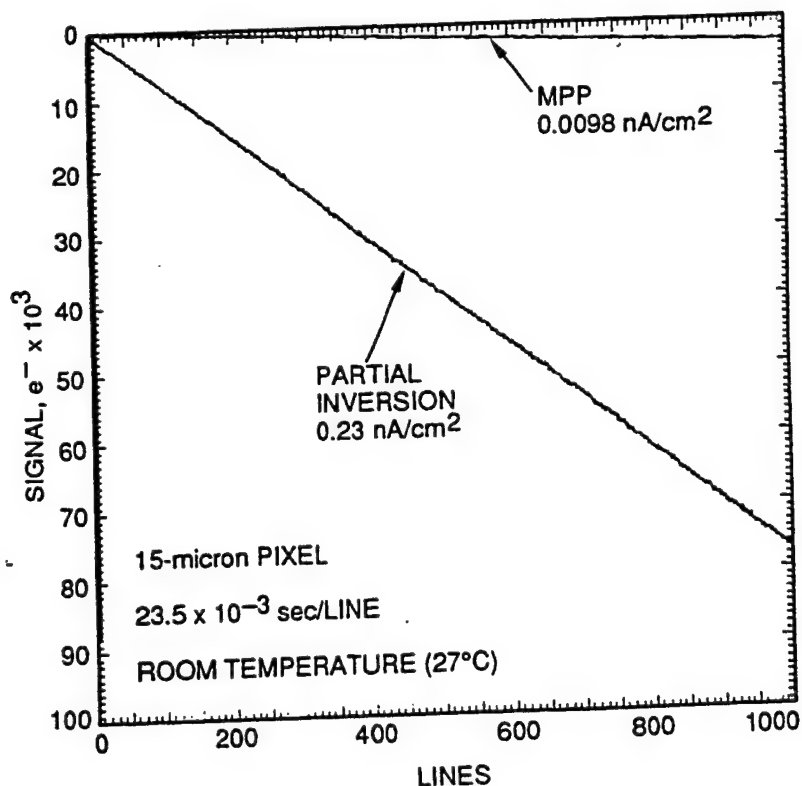


Figure 2b. MPP and PI average dark current build-up after the CCD is quickly erased and then slowly readout at room temperature. MPP dark rate is 20 times less than partially inverted (PI) operation. Dark current for PI clocking is 5 times lower than non inverted (NI) clocking.

Experimental CCD lot runs are sometimes fabricated to find good silicon and achieve low dark current generation. These lots are usually "split" using different types of silicon that are processed identically. Test results have been surprising. For example, silicon purchased from one foundry with the same specifications showed a large variance, 10 pA/cm² to 0.6 nA/cm² operating MPP. Large dark spikes were observed from silicon purchased from another foundry. These spikes generated charge at a rate as high as 10,000 nA/cm². These spikes were later found to be associated with lattice stacking faults that were visible to the naked eye.

As a rule of thumb, bulk dark current for quality silicon is roughly equal to the thickness of the epitaxial layer when expressed in units of pA/cm². For example, a 10-micron epitaxial CCD typically exhibits 10 pA/cm² when fully inverted. Ultra-low dark rates can be achieved by thinning the CCD thereby eliminating substrate and epitaxial interface dark current. For example, MPP dark current for thinned Reticon CCDs exhibit less than 5 pA/cm².

3.2 Charge transfer efficiency

CTE performance varies significantly among CCDs and is also very dependent on silicon material. CTE and MPP dark current performance typically go hand in hand. That is, if bulk dark current is low then CTE performance is also well behaved. This correlation assumes that there are not other CTE impediments that limit performance such as design or process CTE traps.

CTE is specified in pixel transfers at a specified test x-ray signal level. CTE is especially critical to Advanced Camera because ultra-small charge packets will be transferred. For example, the CTE specified for this CCD is 0.999999 for a charge packet size of 1620 e⁻. A deferred charge loss of 7 e⁻ is expected for 4096 pixel transfers.

Bulk traps that happen to lie within the charge transfer channel can trap charge, typically involving a single electron. Figure 3a shows an Fe-55 x-ray (1620 e⁻) response generated by a 800 x 800 Loral Fairchild three-phase CCD that is bulk state limited. We stack, accumulate and display 1024 lines of data into the single trace shown. Note that the single pixel x-ray event line rapidly decreases as more and more charge is fed to trailing pixels (also seen in the plot). Horizontal CTE (HCTE) is very poor (<0.996) whereas vertical CTE (VCTE) is perfect (also measured by x-rays by stacking columns). Data was generated at an operating temperature of -130 C and using a horizontal clock overlap of 1 micro-sec. Operating the CCD at a warmer

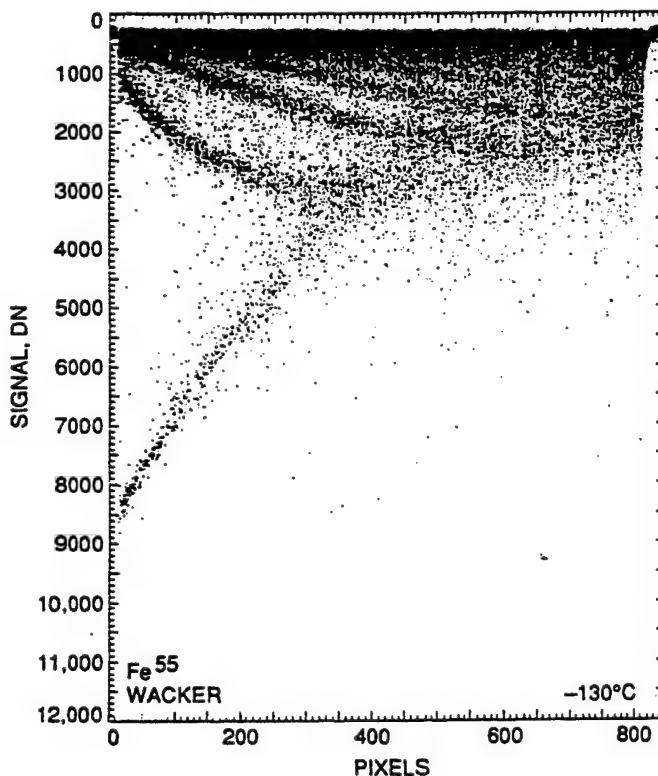


Figure 3a. HCTE x-ray response (Fe-55, 1620 e⁻ / event) for a WF/PC II CCD fabricated on silicon material that exhibits shallow bulk traps. The bulk states result in severe trapping and poor CTE only in the horizontal direction (VCTE is perfect). Deferred charge levels associated with trailing pixels are clearly seen. Note after 300 pixel transfers that more charge is contained in trailing pixels than the target pixel.

temperature or increasing the clock overlap period improves HCTE significantly. For example, Figure 3b and c show horizontal x-ray responses for two different clock overlaps (1 micro-sec compared to 4 micro-sec respectively) at an operating temperature of -110 C. The improvement in CTE is clearly evident when the overlap and temperature is increased. Figure 3d shows HCTE improvement as a function of operating temperature. The response is the composite of nine line stacking x-ray measurements, data taken from the center of the horizontal register. Note that HCTE degrades rapidly for operating temperatures below -80 C. CCDs made on different silicon material in the same lot run did not exhibit the CTE problems over the temperature range or clock overlap period investigated. The difference is striking.

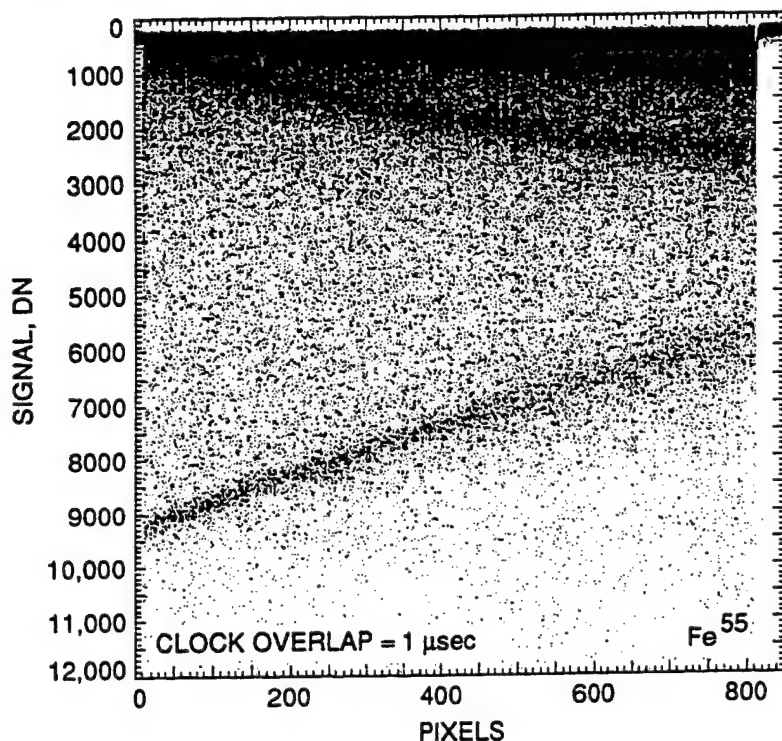
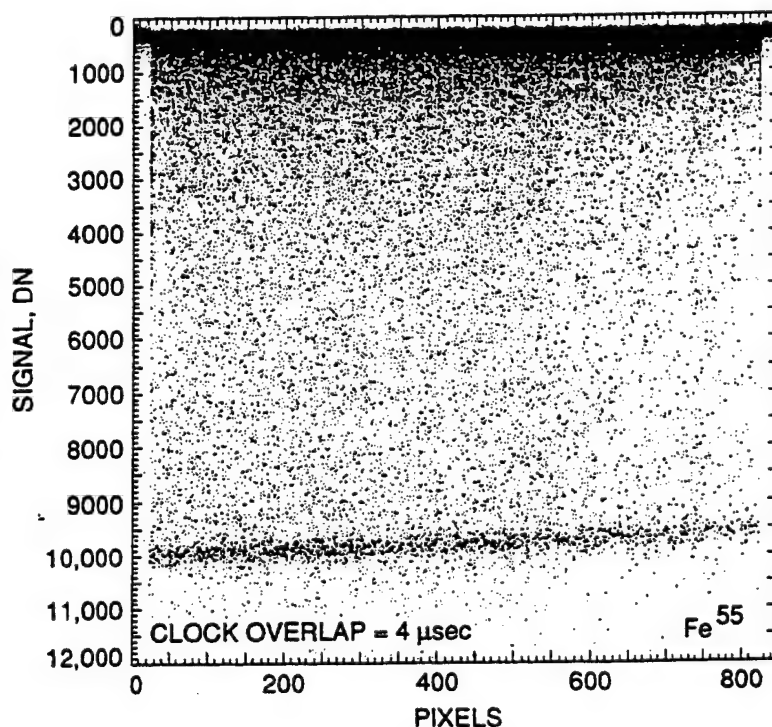


Figure 3b. HCTE x-ray response for the same WF/PC II CCD at a elevated temperature (from -130 C to -110 C). CTE improves significantly because bulk traps detrap charge at a faster rate within the 1-micro-sec clock-overlap period used. Single pixel and deferred charge event lines are clearly seen. Target pixels loose approximately 40 % of charge after 800 pixel transfers.

Figure 3c. HCTE x-ray response when the clock-overlap is increased to 4 micro-sec. Additional transfer time allows charge to escape from traps. CTE performance is nearly perfect. Target pixels at the end of the horizontal register only suffer a 7 % loss making the device usable for most scientific applications..



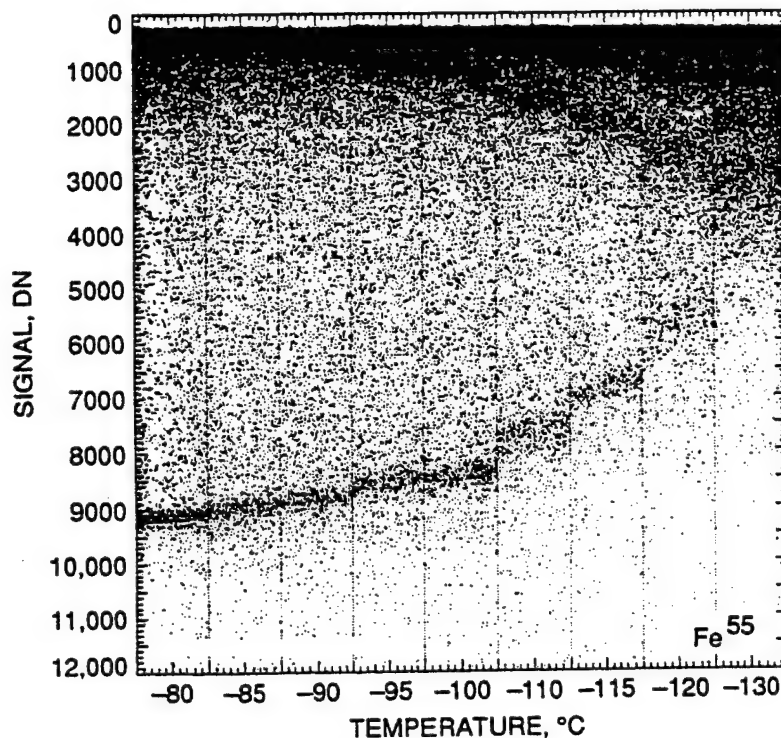


Figure 3d. HCTE x-ray responses generated as a function of operating temperature (data taken from a 400-500 pixel region). Operating temperatures above -80 C exhibit perfect HCTE when the clock-overlap period is greater than 1-micro-sec. These operating specifications are supplied when sensors are delivered for use.

Curiously CTE characteristics for the WF/PC II CCD exhibits completely opposite behavior to the data presented in Figure 3. For example, Figure 4a shows two x-ray stacking column traces for two MPP WF/PC II CCDs built on different silicon wafers (referred to as Type "B" and Type "A"). Data was taken at an operating temperature of -30 C. Note that the VCTE for Type B silicon is poor compared to Type A. HCTE for both sensors is perfect (also opposite to the data presented in Figure 3). Lowering the operating temperature to -90 C for Type A material improves VCTE significantly as shown in Figure 4c. Figure 4d shows similar CTE responses at 0 C. Here CTE varies significantly for Type A material whereas Type B exhibits no temperature dependence. Note also that the MPP dark floor indicated for Type A silicon is much greater compared to Type B (approximately 20 times higher). As mentioned above, CTE and MPP dark current performance go hand-in-hand when performance is bulk state limited. Unfortunately WF/PC II ended up flying silicon Type A material because of schedule problems.

The radical difference between the responses observed in Figures 3 and 4 is due to the emission time constant (T_e) related to bulk traps. For example, the bulk traps that degrade CTE in Figure 3 are relatively shallow traps (trap depth is measured from the conduction band edge). When charge is trapped by a shallow trap it can thermally escape in a short period of time back into the conduction band. CTE can be improved by increasing the operating temperature thereby giving more thermal energy to the carriers to escape the traps faster. Increasing the clock overlap period allows more time for charge to escape the traps. This effect explains why the VCTE is superior to HCTE in Figure 3a. A vertical clock overlap period of 60 micro-sec was used, long enough to allow charge to escape the bulk traps vertically.

When the traps are deep, like those that plague the WF/PC II CCD, completely opposite test results are recorded. Also different solutions are applied to improve CTE. Operating the CCD at a very cold temperature keeps carriers confined to deep traps by removing kT agitation. Once the trap is filled it can remain filled for a long period of time and not trap new charge. As long as there is a small amount of background charge around, the traps will be satisfied resulting in well behaved CTE. Fat-zero charge can come from various sources that are often unknown to the user because levels required are very small (sources of dark charge, spurious charge, sky background, cosmic rays, etc.). Extended images (such as the moon) typically exhibit no CTE difficulties since traps are satisfied by the first charge packets that move through the array (i.e., in this case the first pixels that define the limb of the moon).

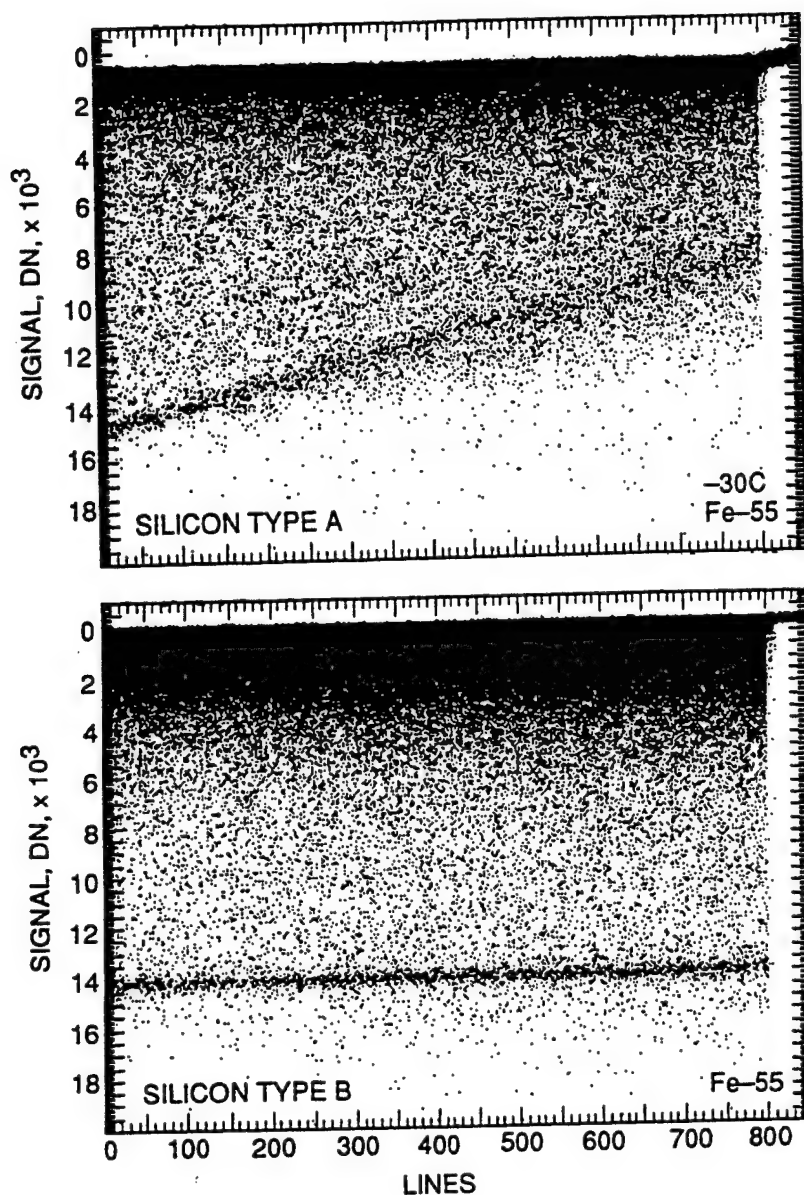


Figure 4a. VCTE responses for two WF/PC II CCDs fabricated on two different types of silicon. Type A silicon shows poor CTE performance because of deep level active bulk states whereas type B exhibits near perfect CTE over all operating temperatures tested. Unfortunately, WF/PC II used Type A silicon because of schedule reasons. However, studying the problem yielded a solution described in text.

When bulk state limited, transferring point images is more difficult especially if widely separated. For example, T_e for the WF/PC II traps are on the order of a thousand seconds when operating at -80 C. At this temperature approximately 10 % of the charge is deferred when measured at the top of the array (assuming a $1620 e^-$ charge packet). This photometry error is based on measuring standard star clusters without fat-zero charge present. Cooling the device to -90 C improves accuracy to 2 % because the traps remain satisfied for a longer period of time because less background charge is required (refer to Figure 4b).

Note that all CTE tests performed above stimulate the CCD with x-rays. X-rays generate an exact known amount of charge making CTE measurements straightforward and absolute. There have been many relative CTE test techniques invented (EPER, dark spike stimulus, optical point source, edge stimulus, electrical charge injection via input diode/gate, etc.,) methods that compare the amplitude of the charge contained in the target pixel to the deferred charge that follows. These techniques usually yield incorrect estimates of CTE by overestimating its true value. For example, WF/PC II star images at -60 C appear sharp with no evidence of deferred charge. Relative CTE methods would have proclaimed perfect CTE for the CCD (in fact photometry error is $> 20\%$). However, as mentioned above, the long emission time constant for the traps spreads deferred charge across the array and is not apparent in the read noise. Only x-rays (or an absolute source such as standard stars) can measure true CTE.

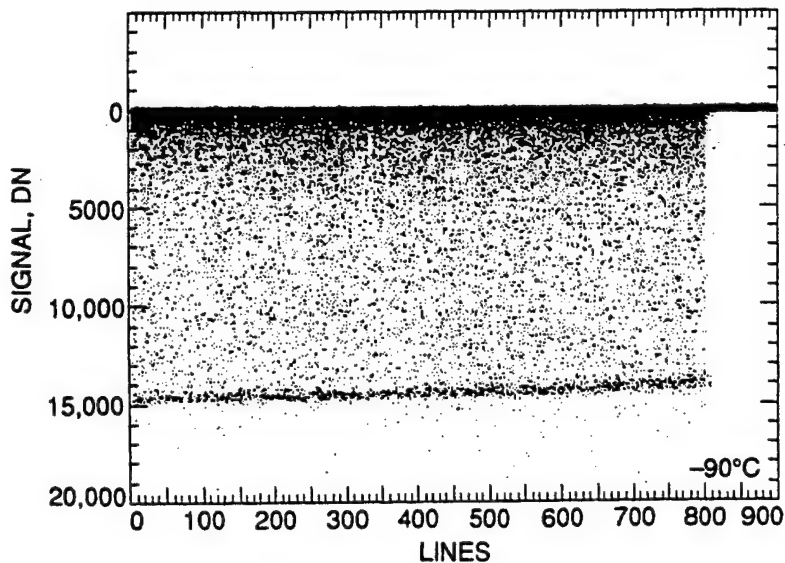
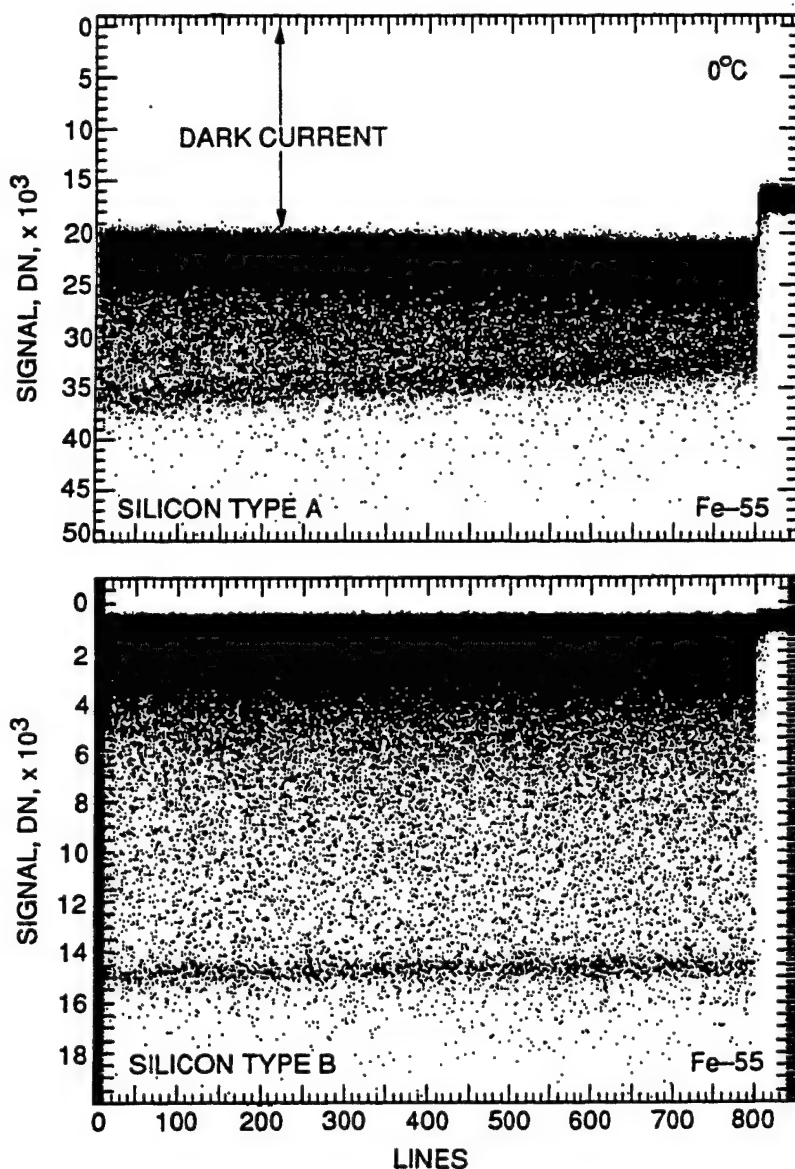


Figure 4b. Type A silicon tested at -90 C exhibiting improved CTE performance (WF/PC II nominal operating temperature on the Hubble Space Telescope). Photometry error at the top of the array (i.e., line 800) is approximately 3 % for a 1620 e- point source of charge. Computer image processing algorithms are used to correct this error below 1%, WF/PCs photometry spec. HCTE performance is perfect for all temperatures tested.

Figure 4c. VCTE responses at 0 C showing differences in MPP dark current generation. Dark current is 20 times higher for Type A silicon compared to type B. CTE and bulk dark current performance are typically related because of bulk states.



Many CCD manufacturers fall victim to relative CTE measurement techniques. We demonstrate the problem in Figure 5a for a 1024 x 1024 CCD with a CTE problem (the sensor was purposely irradiated with neutrons to induce deep level traps, approximately 0.45 eV deep). The 50 x 50 sub-area of pixels shown was taken at the top of the CCD opposite from the lower on-chip amplifier (selected to accentuate VCTE charge problems in the array). After integrating for several seconds charge is transferred down and to the left in the images shown. Horizontal CTE appears perfect because the traps are deep level (i.e., traps are satisfied in the horizontal register because they are constantly being fed). The region selected contains several dark spikes (also induced by the neutrons) and are used as point signal sources (each that generate a constant amount of charge per unit time). The top image exhibits deferred charge tails following each dark spike indicating a CTE problem for the CCD. The lower image shows the same region except that the vertical clocking rate is increased by a factor of 12 (i.e., from 44.5 to 514 lines/sec). Note that the deferred charge tails completely disappear suggesting that CTE has been enhanced by simply clocking the CCD faster. The apparent CTE improvement seen when clocking the CCD faster is investigated further in Figure 5b. A single column trace is presented that contains a large dark spike. The spike and corresponding deferred tail was measured at different array readout rates (23 sec, 19 sec, etc.). Charge is transferred from right to left in the plot. A large deferred tail is seen following the dark spike for a readout time of 23 seconds. As the vertical clock rate is increased the deferred tail becomes smaller and smaller. At a frame time of 1.99 sec the tail is almost eliminated. Although CTE appears to improve (in a relative sense), the amplitude of the dark spike remains roughly the same size independent of clock rate. Since the dark spike generates the same amount of charge during integration deferred charge must also be fixed for each frame time. This behavior indicates

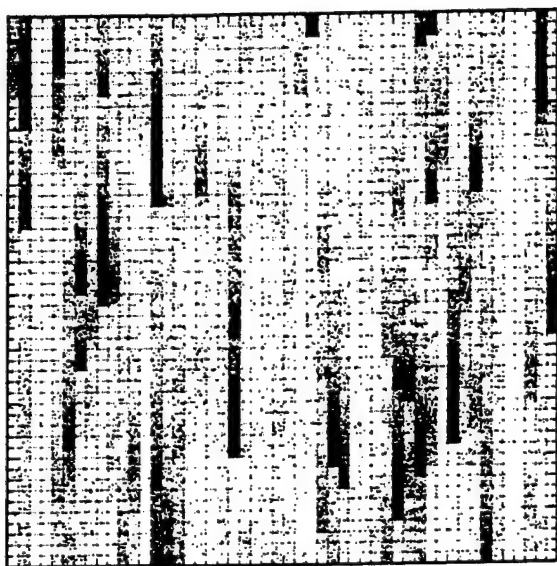
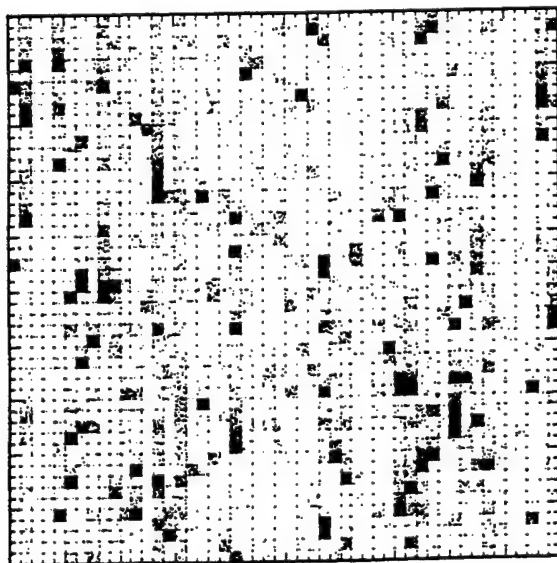


Figure 5a. Two dark spike images generated at two different frame rates. The top image exhibits deferred charge tails whereas the bottom image does not. Comparing the images suggests that CTE performance is better in the lower image.



that deferred charge has spread over more pixels as the rate increases making it more difficult to detect because of the noise floor.

Figure 5c shows the true amplitude of the dark spike by measuring CTE characteristics when transferring charge towards the CCD's upper on-chip amplifier. Clocking the sensor in this manner minimizes trapping with almost no CTE degradation. Comparing the dark spike amplitude in Figure 5b to Figure 5c shows that approximately 2/3 of the charge generated by the dark spike is deferred and is independent of vertical clock rate. CTE performance is very poor at all clocks rates investigated, a conclusion that would not be drawn by only examining the deferred tail.

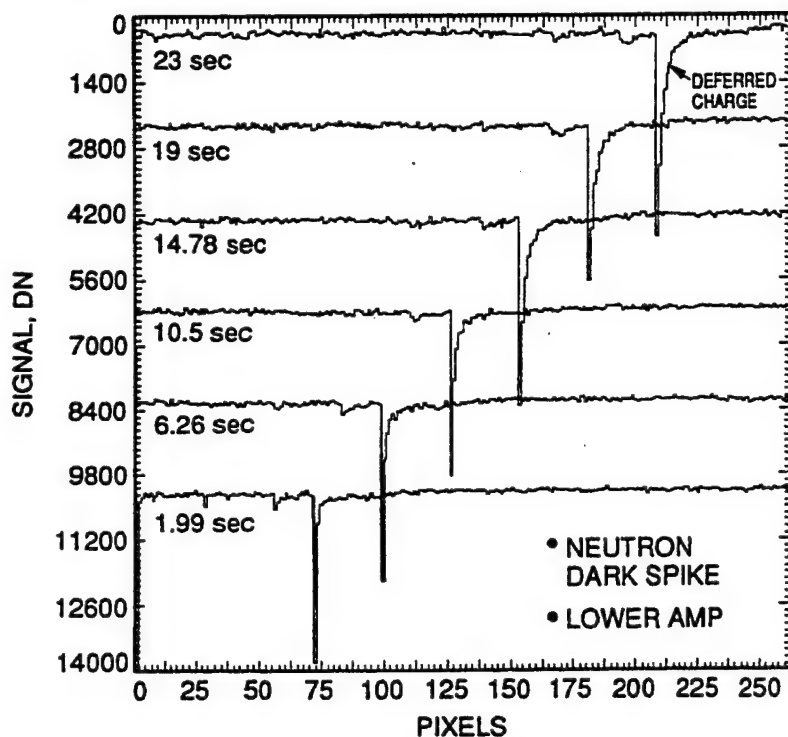
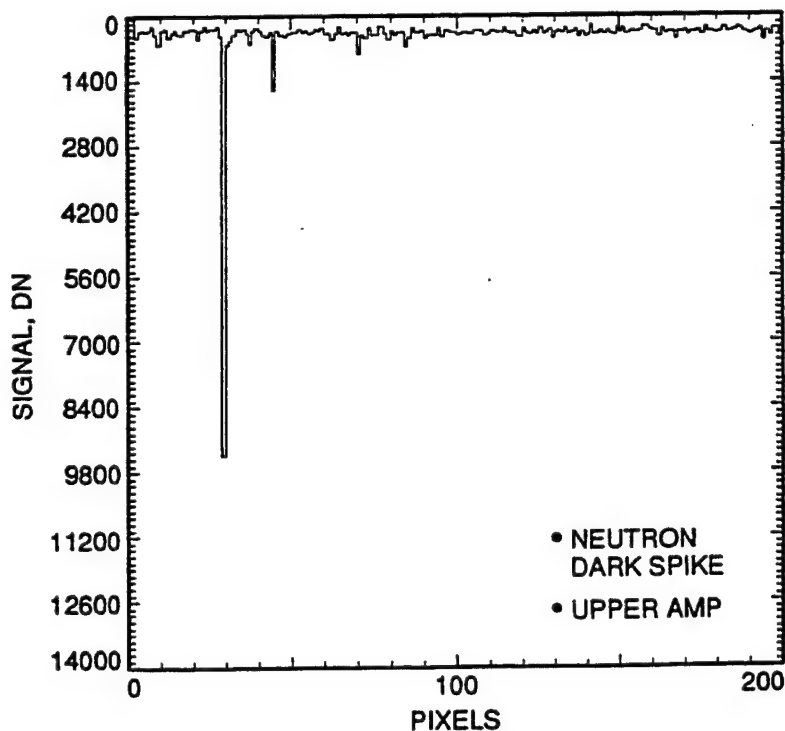


Figure 5b. A single dark spike characterized showing a deferred charge tail. Note when the CCD is clocked faster that the tail disappears implying that CTE improves. However, the dark spike amplitude remains essentially the same indicating that charge is spread over more trailing pixels with rate increase. Hence, speed has no significant effect on absolute CTE performance for the sensor.

Figure 5c. The same dark spike readout using the upper amplifier to minimize the number of vertical transfers. No deferred tail is seen. Also the amplitude of the spike increases by a factor of two compared to lower amplifier readout. Also two small spikes emerge barely seen in Figure 5 b.



The apparent CTE loss, as measured by x-rays (or point sources like stars), is also dependent on the separation between events. If the population of x-ray events on the array is low then the events must go through the array alone. However, if their density is relatively high then events will aid each other in the transfer process. This effect is especially pronounced when no background charge is present. Charge separation effects are demonstrated in Figure 6 where two x-ray column stacking responses are shown. The 1024 x 1024 CCD tested is partially damaged with protons. Half of the array is shielded and not damaged. The CCD is exposed to x-rays and then quickly clocked vertically (600 micro-sec/line) reading out approximately 512 lines. On the average there are 100 x-ray events per line. The data associated with these pixels is discarded by the computer. The remaining half of the array is then clocked slowly (20 ms/line) and saved by the computer. Vertical stacking plots are then generated for the damaged and shielded regions as displayed in Figure 6a. Note that the damaged region (the tilted response) begins at the ideal level for the shielded region (the flat response). This indicates that during high speed clocking that most x-ray events are not trapped because separation between events decreases compared to slow-scan readout which exhibits a tilt. Figure 6b shows the results of a similar experiment where the separation between events is reduced by a factor of 100 (i.e., 1 event/line). Note that the response for the damaged region does not begin at the ideal level indicating that charge is trapped more often during high speed readout because events are widely separated.

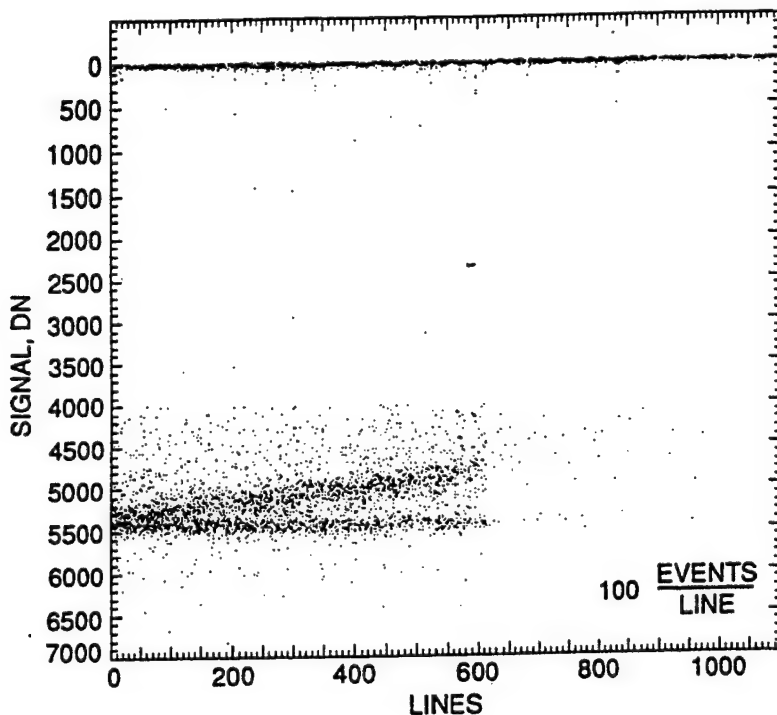


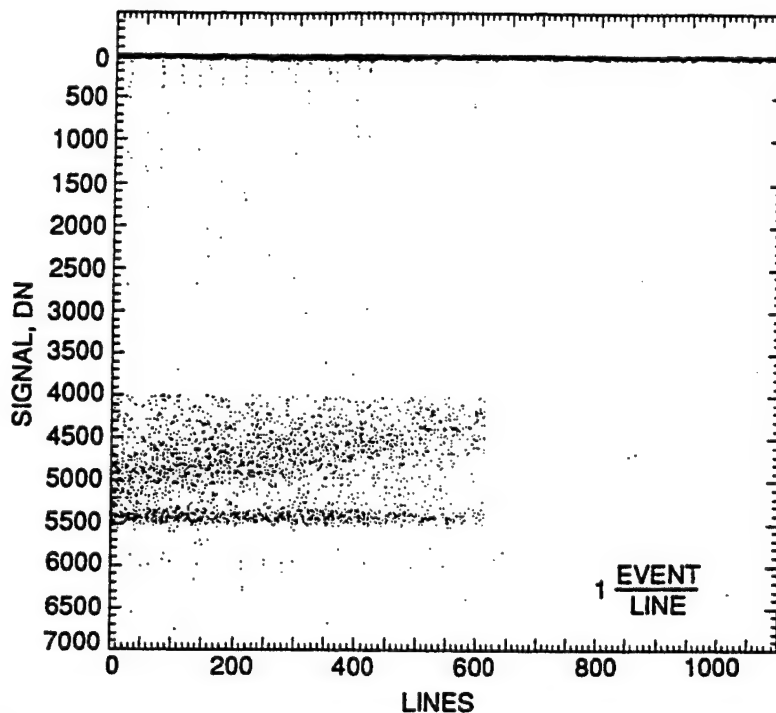
Figure 6a. VCTE responses taken from two subarrays - proton damaged and shielded regions. The damaged region exhibits good CTE when clocked at a fast rate (600 micro-sec/line). The single pixel event line for the region begins at the ideal level. Under these conditions the average time between x-ray events during readout is $< T_e$, hence events aid each other in the transfer process. Reading the CCD slow-scan (20 ms/line) causes the CTE to degrade in the damaged region resulting in the tilted response shown. The average time between events in this case is $> T_e$ and most events must go through the array alone. The shielded region shows good CTE.

The capture time constant (T_c) associated with deep traps also influences CTE results. Theory claims that T_c is very short (less than 1 micro-sec) and electrons are trapped instantaneously when transferred into a potential well. Therefore, it is generally assumed that $T_c \ll T_e$. This is true for the most part, however, electrons at the bottom of a well can be thermally agitated forcing them to move up the sides of a potential well and find new traps. The more time given to the electron the greater the probability it will find a trap. This effectively increases T_c for some traps. Hence, for this reason it is best to clock the CCD as fast as possible to minimize this trapping effect. This requirement only applies to deep traps.

For some CCDs charge trapping can become so severe that the device must be light flooded to fill in the traps and then erased quickly before an image is taken (the Galileo CCD - a 800 x 800 CCD in route to Jupiter, and the Cassini CCD are used in this manner because of radiation induced bulk traps). However, for this technique to work properly the operating temperature must be sufficiently low to keep the traps filled during integration and readout (e.g., Galileo operates at -120 C).

Therefore, T_c , T_e , operating temperature, clock overlap, clock rate, background charge, charge distribution, and charge packet size are important factors that influence CTE when bulk states exist (there are others). Ideally, CCDs should be built on quality silicon to avoid CTE and dark current problems discussed above. However, no matter how good the material some bulk states

Figure 6b. VCTE response when the x-ray event density is reduced by a factor of 100. This effectively increases the time between events. Both fast-scan and slow-scan exhibit degraded CTE in the damaged region.



will always exist. As CCDs become larger (such as the 4096 x 4096 Sandbox CCDs) and performance requirements get tighter, these variables must be completely understood and controlled by the user if optimum CTE performance is to be realized.

Unfortunately no method has been devised to determine the quality of silicon before a lot of CCDs is run. The only successful approach has been to purchase many silicon wafers and include some of them in ongoing lot runs to be evaluated later over a wide temperature and clocking range (to cover all applications intended). If the silicon tests good then the wafers can be used in future lots.

The technique of "pocket pumping" is one method often used in our lab to characterize silicon material once a CCD has been fabricated. This method allows one to locate and count the number of active bulk traps that influence CTE performance. For example, Figure 7a shows a pocket pumping map generated for quality silicon material. The density of traps measured is 0.002 single electron active traps per pixel based on a 1600 e⁻ flat field signal and an operating temperature of -100 C (trap density changes when these variables are changed for reasons given above). CTE for this CCD is exceptionally high yielding a CTE > 0.999999 under the conditions tested here. Figures 7b and c show raw pocket pumping responses. Figure 7b was taken at an operating temperature of -90 C whereas Figure 7c was generated at -120 C. Note that some traps freeze-out at -120 C (i.e., T_c for these traps become longer than the clock overlap employed in the pocket pumping experiment).

From the discussion above dark current generation and CTE performance can be a hit and miss game. Fortunately the Sandbox CCDs have selected quality silicon based on previous lot builds. Therefore, low dark current (< 1 nA/cm² noninverted and < 20 pA/cm² MPP) and well behaved CTE performance are expected. However, some operational tricks, such as those demonstrated above, may be required.

3.3 Well capacity

Full well is an important characteristic to the Sandbox CCDs especially for the 9-micron pixel devices where dynamic range is at a premium. This section discusses how both the user and manufacturer optimize full well performance for small pixel devices.

3.3.1 Surface and bloomed full well operation

The gates of CCD can be clocked in three different ways. We refer to these clocking modes as noninverted (NI), partially

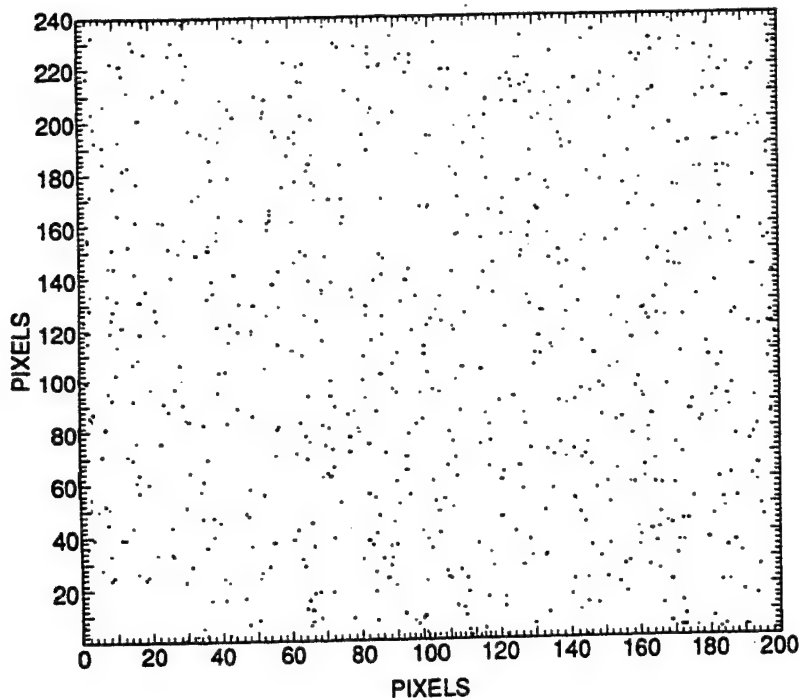
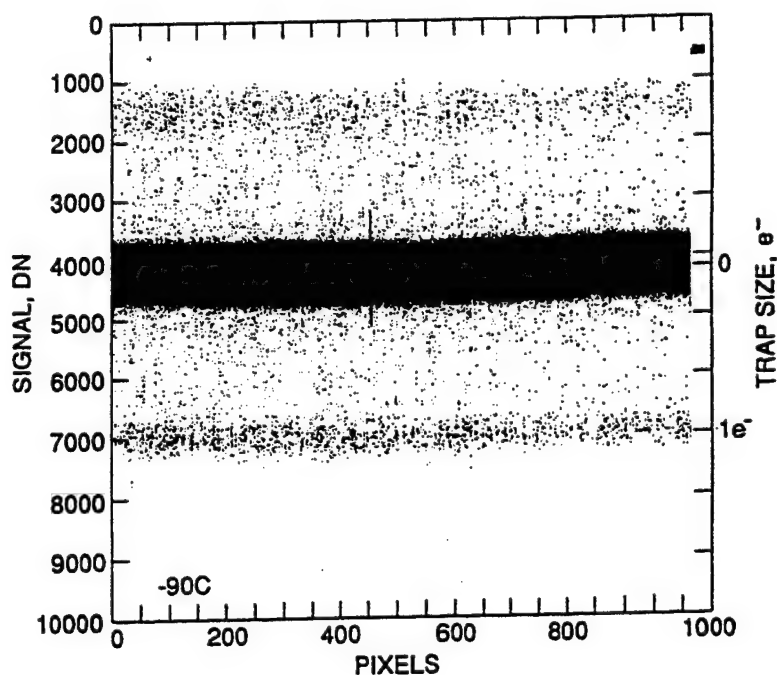


Figure 7a. Array pocket pumping trap map showing individual 1 e- array traps. A 240 x 200 pixel region is shown. A trap density of 0.002 traps/pixel is measured for the operating temperature(-90 C), vertical clock-overlap (60 micro-sec), and signal level (4000e-) employed. CTE for the CCD is exceptional, > 0.999999 as measured by x-rays. Decreasing operating temperature or clocking the CCD faster effectively reduces trap count further improving CTE. Increasing the signal level increases trap density because the charge packets occupy more volume in the channel.

Figure 7b. A raw pocket pumping response showing individual 1 e- traps at an operating temperature of -90 C. Traps less than 1 e- are statistical, trapping and releasing charge randomly.



inverted (PI), and multi-pinned phase (MPP). The user can select the best mode for the application. However, it is often best to design a camera system so that all three modes can be utilized on command. NI operation implies that the clocks to the CCD never swing low enough to invert the signal channel (e.g., -3 to +5 V). PI means that clocks invert as they switch (e.g., -8 to +5 V). For a three phase CCD one phase must be high at all times to keep charge confined to a pixel (hence the name partially inverted). MPP operation is where all clocks are normally inverted most of the time except during line transfer (e.g., -8 V). Each mode of operation has a profound effect on CCD performance. Well Capacity, dark current generation, residual image, anti-blooming, radiation damage tolerance, charge transfer, and pixel nonuniformity are some CCD parameters influenced by these clocking modes. We only examine well capacity here.

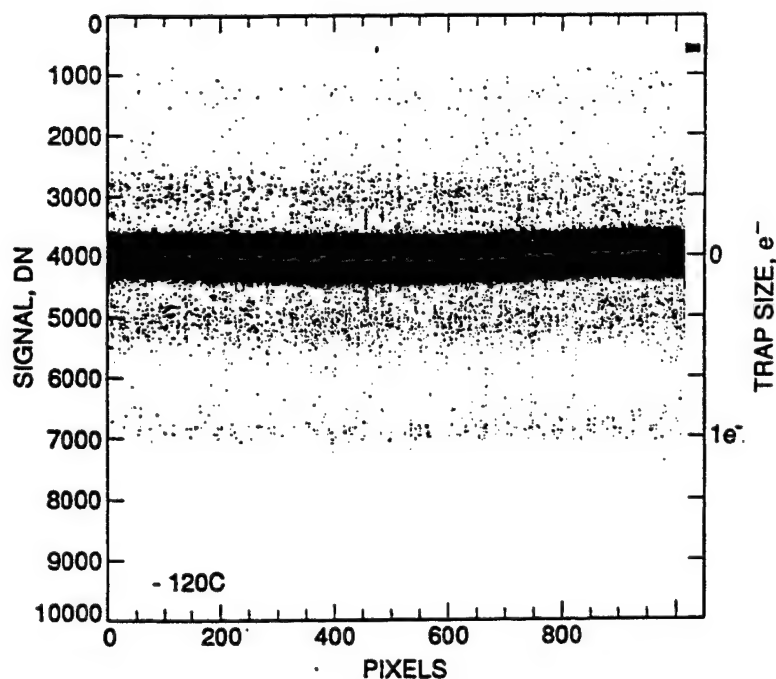


Figure 7c. Pocket pumping response taken at an operating temperature of -120 C. Note that some traps begin to freeze-out at this temperature (i.e., traps trap and hold charge over pocket pumping cycle). Other shallower traps are not affected by the temperature change indicating that at least two different bulk states are involved (possibly phosphorus and oxygen vacancies). X-ray measurement show that CTE improves when the temperature is lowered.

Optimum full well is achieved when the CCD is clocked into inversion. The increase in well capacity under inverted clocking can be significant compared to noninverted operation. For example, full well for the first generation Texas Instruments Space Telescope CCDs (WF/PC I) could have been increased by a factor of three over what was originally flown if clocked PI. Unfortunately the flight units were clocked NI (from 0 to 7 V) yielding only 30,000 e⁻ well capacity. One year after these CCDs were placed into orbit flight spare units were clocked into inversion for the first time (approximately -8 V) resulting in a full well of 90,000 e⁻. Additional capacity was achieved when the collecting phases were driven to +4 V yielding 100,000 e⁻ overall. The apparent full well increase was not clear when the tests were performed for results were obtained by trail and error methods. We now, several years later, understand how to optimize well capacity for the CCD. A brief discussion on full well theory follows.

Well capacity for a pixel is defined when charge either blooms over a barrier phase into adjacent pixels or when charge in a potential well begins to interact with the surface at the Si-SiO₂ interface. Blooming is reduced by biasing barrier phases to the lowest channel potential possible (i.e., inversion). This condition allows for maximum charge collection and the greatest potential difference between the collecting and barrier phases.

A potential well collapses when charge is collected. Blooming will result if the channel potential for the collecting phase equals the potential of the barrier phase. We refer to this condition as "bloomed full well" (BFW). Also as charge collects the potential maximum also moves towards the surface. Therefore, it is possible for the potential maximum to first reach the surface before the onset of blooming. If this happens we say the CCD has reached the "surface full well" (SFW) state. Which comes first, BFW or SFW, depends solely upon the positive clock level of the collecting phase. Optimum capacity occurs when BFW and SFW occur simultaneously, a very important bias state that user must find for optimum performance.

Figure 8 plots full well as a function of positive drive for a three-phase CCD. BFW and SFW regimes are indicated in the figure. Data was generated in the following manner. A 100 x 100 pixel region on the array is saturated with light. If the CCD is biased in the BFW regime then charge will run up and down columns from the exposed region. After the shutter is closed extra time (a few seconds) is given to allow this process to reach equilibrium (blooming is a time dependent process). The resultant signal level is measured in the area representing a BFW data point. To detect SFW another procedure is used. After integration, with the shutter closed, we clock phases 2 and 3 while leaving phase 3 inverted as a barrier preventing charge from transferring. If the CCD is SFW limited charge will get trapped at the Si-SiO₂ interface under phases 1 and 2. Each time these phases are switched into inversion holes recombine with these trapped electrons. Eventually, if the device is clocked long enough (a couple of seconds at 2000 Hz is satisfactory), the charge level will reach equilibrium at SFW (this same technique is also

used to prevent blooming when the shutter is open). Signal is measured in the region representing a SFW data point. These two measurement techniques are performed in series to generate the full well curve shown in Figure 8.

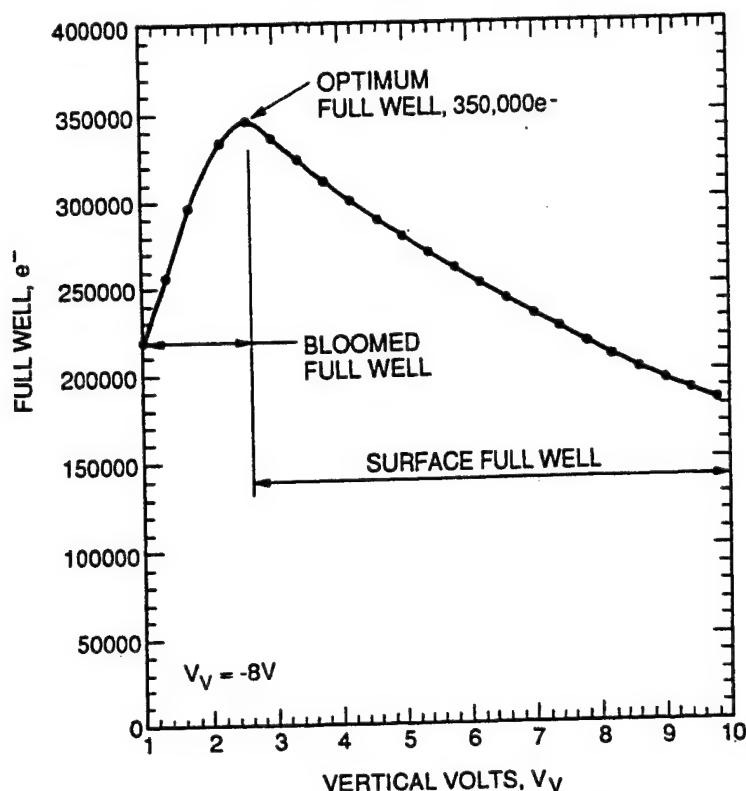


Figure 8. Full well characteristics as a function of positive drive using PI clocking. BFW and SFW regimes are indicated. Optimum full well occurs where charge interacts at the surface simultaneously when charge blooms over barrier phases. Data was generated by a 1024 x 1024, 18-micron pixel CCD. Optimum charge capacity for the CCD is 350,000 e^- .

Note in Figure 8 that optimum well capacity occurs when the collecting phase is driven to 2.5 V yielding 350,000 e^- . At this point BFW = SFW. Increasing the gate voltage above optimum moves the channel potential closer to the surface reducing SFW. At some very high gate voltage (not shown on plot) the potential maximum reaches the surface at which point SFW = 0. BFW is extraordinarily high under this clocked condition since the potential difference between the collecting and barrier phases is several volts. When clocked in the BFW regime the sensor will never run surface channel. Instead charge will bloom up and down the channel. BFW is reduced to zero when the potential of collecting and barrier phases are equivalent (i.e., when all are driven into inversion).

3.3.2 Full well process and design considerations

The discussion above demonstrates how the user optimizes full well performance by the manner in which the CCD is clocked (additional discussions are given below). The CCD manufacturer also plays an important role in maximizing well capacity (Ref. 3). Different design and process methods are employed to accomplish this. The depth and doping concentration of the n-channel are the two main process methods used to control full well. SFW and BFW both increase with channel doping. The amount of charge that can be collected in a potential well is proportional to the number of ionized phosphorus atoms in the n-layer. If the channel is lightly doped then the potential well will collapse faster as electrons are collected (limiting BFW). Also the rate at which the potential maximum moves towards the surface is faster for lightly doped channels (limiting SFW).

CCD manufacturers attempt to use maximum channel doping to increase charge capacity. However, there is a practical limit to the doping concentration that can be employed. For example, as the doping level increases higher gate clocks are required to control the collecting and barrier phases (both that stress the gate insulator). Channel depletion also requires higher bias to the channel (i.e., V_{REF}). However, the main limitation is associated with the electric fields generated in the middle of the channel near the surface where the fields are strongest. If the doping is excessive, high fields will generate dark spikes. In fact, fields can be fabricated where the CCD avalanches, saturating the CCD permanently. High internal fields also make the CCD vulnerable to radiation damage problems. For example, dark spikes are generated when the silicon lattice is damaged and high fields are present (i.e., field assisted dark current emission). Radiation sources, such as protons, induce lattice damage which

in turn leads to dark spike problems (refer to Figure 5).

Channel doping also influences characteristics of the on-chip amplifier. The performance of this transistor will degrade if channel doping is too heavy. This in turn results in higher read noise and nonlinearity for the device. To circumvent this problem manufacturers dope the amplifier and array independently using different reticles. The Sand Box CCDs are fabricated in this fashion where doping concentration of the array will be heavier than the amplifier to optimize charge capacity ($1.9 \times 10^{16} \text{ cm}^{-3}$ and $1.6 \times 10^{16} \text{ cm}^{-3}$ total dose respectively).

Although not as influential as channel doping, SFW conditions improve when the n-channel is made deeper. This is accomplished by an extended, high temperature cycle that drives phosphorus into the epitaxial layer. A deeper channel moves signal charge away from surface improving CTE (suppresses traps) and SFW characteristics.

Channel doping for three-phase buried channel CCDs is limited to about $2.5 \times 10^{16} \text{ cm}^{-3}$ (assuming box-like distribution - surface doping is approximately 2x higher) before dark spikes and clock drive become uncontrolled. The WF/PC II and Cassini CCDs employed channel doping of $1.6 \times 10^{16} \text{ cm}^{-3}$ and a 45 minute/1075 C channel drive. Well capacity for the CCD, when clocked PI, is approximately 4200 e- per square micron of active channel (i.e., less channel stops). Both the Cassini and WF/PC II CCDs were designed with 2-micron channel stops which encroach into the channel approximately 4-microns after processing. The Cassini CCD is a 12-micron pixel and the Space Telescope a 15-micron pixel. Hence, full well for the Cassini and WF/PC II CCDs achieve a full well of 134,000 and 231,000 e- respectively.

It is interesting to compare full well performance for the original WP/PC I CCD to the new WF/PC II CCD fabricated 15 years later. Recall from above discussions that well capacity for the WF/PC I was only 30,000 e-, a factor of 7.7 times smaller than WF/PC II for the same size pixel. The full well increase has come from design, process and clocking improvements made over the years.

3.3.3 MPP full well

When the first MPP CCDs were fabricated full well performance capacity was far from optimum. For maximum well capacity the doping concentration (usually boron) beneath the MPP phase (typically phase 3 for a 3-phase CCD) must be carefully selected. Without the MPP implant full well is zero since the channel potentials for all three phases are equivalent when inverted. Adding boron compensates phosphorus atoms that define the n-buried channel. Boron causes the potential under phase 3 to decrease relative to phases 1 and 2 allowing charge to collect. Well capacity for a MPP CCD therefore initially increases with boron concentration. However, doping phase 3 compensates the phosphorus atoms that define the buried channel. This in turn causes the potential maximum of the channel to move closer towards the surface lowering SFW capacity under this phase. Optimum full well is realized when the charge capacity under phases 1 and 2 equals the charge capacity of phase 3. Eventually if all phosphorus atoms are compensated phase 3 operates completely surface channel yielding SFW = 0 (indicated in figure).

Figure 9a illustrates, with two curves, the characteristics described above by plotting full well as a function of boron concentration. The first curve plots BFW characteristics for phases 1-2 and the second curve for SFW for phase 3. Charge capacity for the CCD is determined by the lowest full well curve shown. Optimum full well occurs when SFW for phase 3 equals BFW for phases 1 and 2, a condition that manufacturers attempt to achieve when fabricating MPP CCDs. Channel doping and drive also influence the optimum MPP implant employed, subjects discussed in considerable detail in Reference 2.

Figure 9b plots full well as a function of gate voltage (phase 3) for five MPP CCDs (shown dashed in Figure 9a). Consider sensor #1 first. When $V_g = -2 \text{ V}$ the full well is zero because the channel potentials are equivalent for all phases (i.e., MPP phase 3 is high at -2 V and phases 1-2 are low and pinned at -8 V). This clocking condition will occur when readout commences. Increasing the gate voltage results in greater charge capacity for phase 3 again during readout. At $V_g = 0 \text{ V}$ phase 3 holds more charge during readout than phases 1-2 during integration. Full well becomes limited by these phases exhibiting the flat response with increasing V_g as shown. SFW for phase 3 will eventually occur at $V_g = 8 \text{ V}$. At this point full well decreases as phase 3 is driven deeper into the SFW regime.

Now consider device #2 which employs more MPP implant. First note that full well = 0 when $V_g = -1 \text{ V}$, a shift of 1 V compared to device #1 because of the additional boron implant. Increasing the doping increases full well for phases 1-2 because

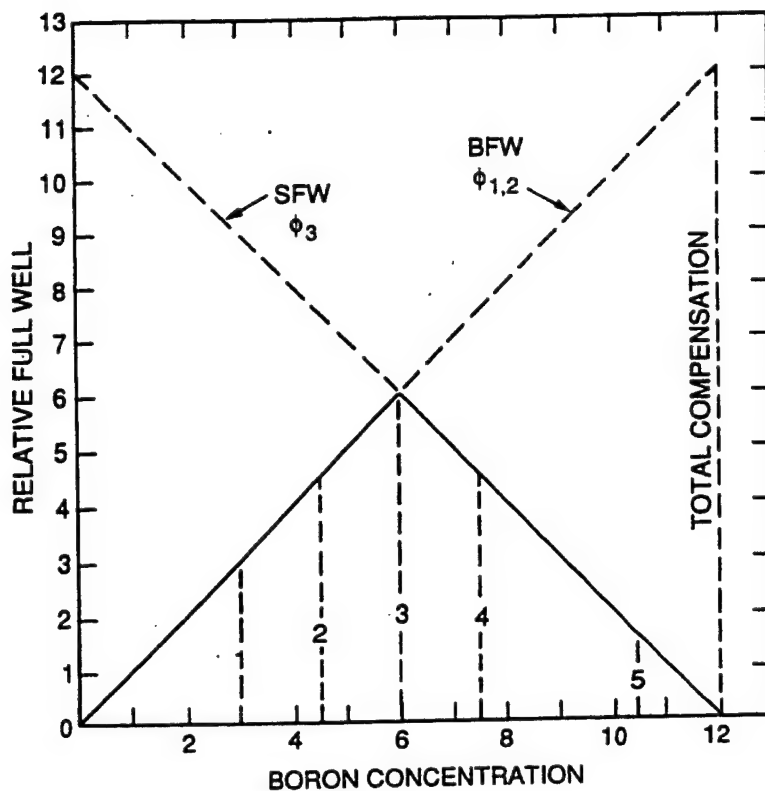
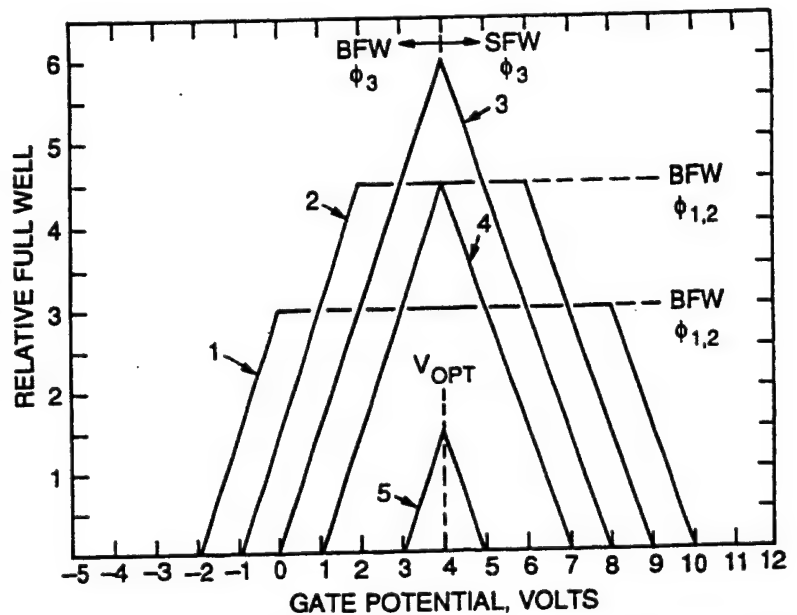


Figure 9a. Full well for a MPP CCD as a function of boron implant when clocked MPP. Optimum doping is achieved when the charge capacity under phases 1 and 2 during integration equals the charge capacity under phase 3 during readout. Doping beyond optimum results in surface channel operation for phase 3 and lower full well performance.

Figure 9b. Charge capacity as a function of gate voltage for phase 3 for five different MPP implants. The MPP implant for device 3 has been optimized to yield the greatest full well possible. Devices 1 and 2 are limited by the amount of charge that can be stored in phases 1 and 2 during integration. Devices 4 and 5, which have an implant greater than optimum, are limited by phase 3 and SFW during readout.



the barrier height increases under integration. SFW for phase 3 is offset by 1 V since more boron reduces SFW characteristics. Therefore, both phase 3 BFW and SFW curves shift towards V_{opt} by the same amount. The implant used for device #3 is optimum since $BFW = SFW$ for phase 3. Devices #4 and #5 employ an implant dose that yields a full well level less than optimum (limited by SFW for phase 3).

3.3.4 MPP and PI clocking

Early MPP Cassini CCDs used a boron implant less than optimum, approximately $4.5 \times 10^{15} \text{ cm}^{-3}$ (optimum is $6 \times 10^{15} \text{ cm}^{-3}$ at 60 keV). MPP full well characteristics were therefore modest achieving only 30,000 e⁻ (compared to 80,000 e⁻ for an

optimum implant). However, when the CCD is clocked PI a significant increase in full well is measured, approximately 110,000 e⁻. However, this full well is less than a non MPP CCD clocked PI where 134,000 e⁻ is achieved. This behavior is typical for MPP CCDs. MPP well capacity is always less than non MPP devices when clocked PI. This is because the boron implant reduces SFW and BFW characteristics for phase 3.

Figure 10 illustrates full well characteristics for a non MPP CCD (sensor #1) and three MPP CCDs (sensors #2, #3, and #4). Sensor #2 is a MPP CCD that incorporates an implant less than optimum. As discussed above, full well for the CCD is limited by the amount of charge that can be stored under phases 1-2 during integration when all phases are inverted. However, when clocked PI these phases are biased high allowing for much greater charge capacity. Under these conditions full well is determined by SFW and BFW characteristics for phase 3. The increase in full well is significant if the MPP boron dose is low (dotted line in Figure 10). The MPP implant for sensor #3 is optimum. PI and MPP clocking yield equivalent full wells. Sensor #4 dose is greater than optimum. Full well is dependent on SFW characteristics for phase 3. Hence, there is no advantage to implanting a CCD beyond optimum.

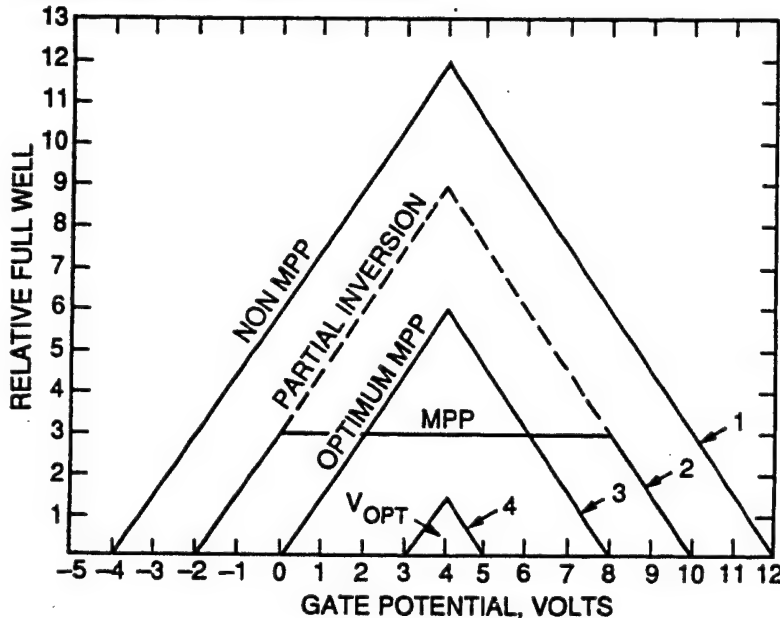


Figure 10. Well capacity as function of clock potential for a non MPP clocked PI (device #1) and three MPP CCDs (devices #2-4). Early MPP CCDs were processed similar to device #2 where well capacity is greater when clocked PI than when clocked MPP as shown. For the optimum MPP implant, PI and MPP clocking yield the same well capacity and full well curve. An implant greater than optimum also produces the same full well response but at a much reduced level. Note that all curves peak at V_{OPT} . This is because the MPP implant shifts the potential for BFW and SFW characteristics by the same amount.

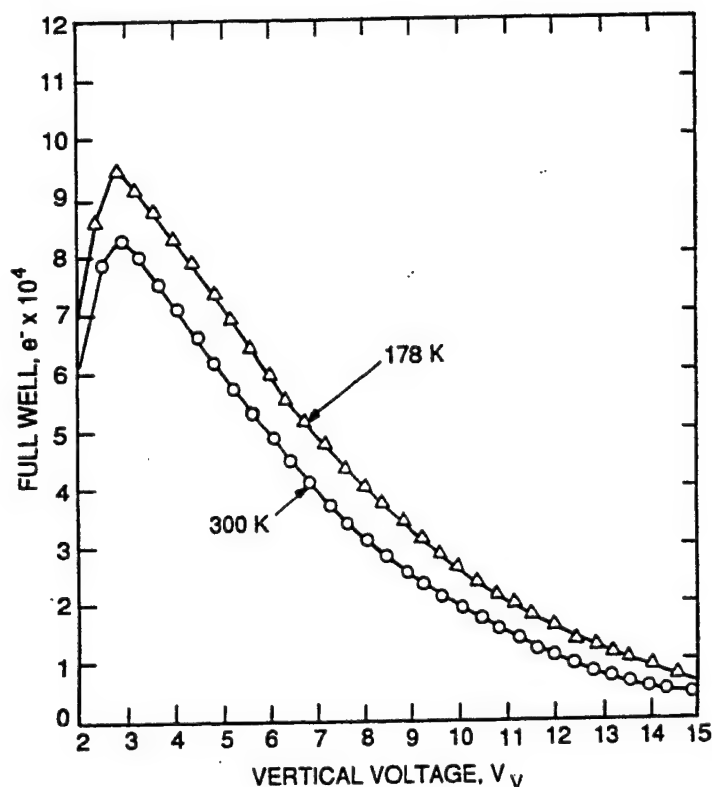
The Cassini MPP CCD exhibits optimum MPP full well performance. Figure 11 plots full well characteristics for the device as a function of gate voltage for two operating temperatures. Data was generated using MPP clocking, however, the same curves are generated when PI clocking is used for reasons given above. Note that full well increases when operating temperature is reduced. Full well in general is a temperature dependent variable. Strictly speaking, the onset of full well actually occurs before the channel potential equals that of the surface (SFW) or barrier phase (BFW). As charge collects in a potential well it naturally wants to come out. Two primary mechanisms are responsible for this action. First a repulsion force is generated between electrons that are collected. This field attempts to force carriers out of the well. Second, electrons also gain energy thermally (i.e., lattice vibration). The energy acquired is not the same for each electron but is described by Boltzmann's statistics. If the energy kick is large enough the electron can bloom over a barrier field. Electrons can acquire as much as 100 kT (2.5 eV) if sufficient time is given to acquire such energy statistically (hence full well is also time dependent). A similar process occurs when a CCD goes SFW. Here charge that collects near the surface (approximately 1000 Å) can thermally hop into the Si-SiO₂ interface region where the first signs of SFW appear.

For some applications it may be desirable to only introduce a small MPP implant barrier to preserve full well performance but still achieve the advantages of full inversion (although MPP full well will be modest). The user can command the camera to clock the CCD either MPP or PI. For low dark current generation and low signal levels MPP clocking would be advantageous. For high signals, requiring greater dynamic range, PI clocking could then be commanded.

The Sandbox CCDs will not include the MPP option for four reasons. First, dark current will not be an important factor since the CCDs will either be sufficiently cooled (e.g., -90 C operation for the Advanced Camera CCD) or readout quickly (e.g.,

30 frames/sec for the Cinema CCD), both methods that suppress dark current to acceptable levels. Second, dynamic range is important especially for the 9-micron pixel devices. All CCDs will be clocked PI to achieve maximum well capacity. As described above, full well is sacrificed when MPP is implemented. Third, although low dark current generation is desirable for Advanced Camera and Pluto Flyby CCDs we anticipate (based on WF/PC II experiences) that radiation dark spikes will be generated when these CCDs go into space. Dark spikes will therefore limit dark current generation requiring cooling below -90 C. Fourth, MPP operation can severely limit how fast a CCD can be readout, an effect described below. This last factor is important to most Sandbox CCDs because they will be clocked at very fast rates in order to perform the frame transfer operation.

Figure 11. Full characteristics for the Cassini MPP CCD. This sensor was processed with an optimum MPP implant. Full well characteristics are the same whether the sensor is PI or MPP clocked. Well capacity is dependent on operating temperature as explained in text. At room temperature full well is 82,000 e⁻ for the 12-micron, 3-phase pixel.



3.3.4 ESD protection networks

The discussion above shows the importance of driving the CCD with bipolar drivers to achieve optimum full well performance. Commercial CCDs often use protection diodes in parallel with clock lines to prevent electro-static damage (ESD). However, driving these diodes negative (relative to substrate) causes them to conduct resulting in a short circuit condition. This results in severe luminescence on the array (LED effect), not to mention the potential harm to the drivers and diodes. Also flight CCDs must pass a DC probe short test. For example, JPL flight CCDs require that poly to poly or substrate shorts < 1 M ohm (ideally < 10 M). This impedance test cannot be performed if diodes are present because the impedance of the diode would limit such measurements. Hence, for these two reasons ESD diodes are not employed for the Sandbox CCDs.

It has been often observed that when an ESD short does occur it is located at the edge of the array where poly lines drop from the field oxide region down into the active region (i.e., the interface of the p⁺ moat). It is this region where oxide fields are greatest when a high voltage spike is applied. An ESD short is typically not a dead short but a relatively high impedance leakage path where free carriers can be injected into the array. A bright column blemish or luminescence are signatures of a ESD short if the problem is associated with the vertical registers. Clock drivers normally can drive shorts of this type allowing the rest of the array to be readout normally.

The gates of the Sandbox CCDs are somewhat protected. The protection network consists of a small p⁺ moat in series with the aluminum bus line located near the bond pad. If a high voltage spike does appear damage will likely be induced in the network and not on the array. Hence, charge injection and luminescence will be minimized. Diffusions (e.g., V_{REF} and V_{dd}) are not

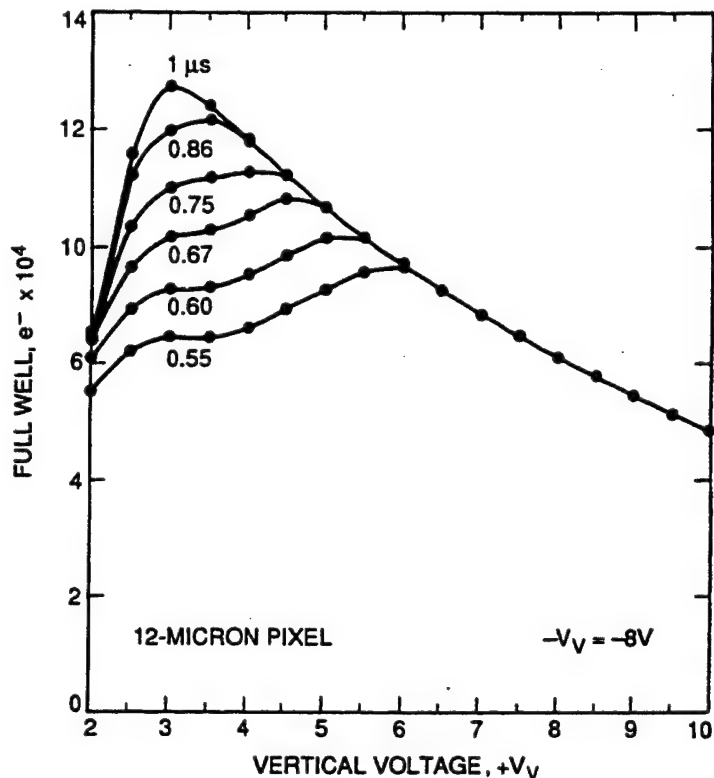
protected since they are not ESD sensitive (relatively speaking).

4. HIGH SPEED CLOCKING

The Pluto Flyby, Adapt II and Cinema CCDs will be clocked at a very high scan rate to perform frame transfer. The transfer time into the storage section will be performed 100 times faster than the longest integration time to suppress smearing. For example, the integration time for the Cinema CCD is approximately 1/30 of a second. Therefore, a frame transfer time on the order of 0.33 ms is required, a line transfer rate of 0.32 micro-sec to transfer 1024 lines.

A number of interesting limitations develop in the CCD when lines are transferred at this rate. First, well capacity decreases as transfer rate increases. Figure 12 plots well capacity for a 1024 x 1024 Cassini CCD as a function of clock potential ($+V_V$) with PI clocking employed. A family of curves are presented for different line transfer periods (the shortest equal to 0.55 micro-sec). Data is collected in the following manner. The CCD is first erased of all charge. Then a 100 x 100 pixel region is illuminated at the top of the array (coordinates approximately 750 x 750). The illumination level is slightly greater than optimum full well. The CCD is clocked vertically 512 times with the line time indicated and data discarded. The remaining 512 lines, still in the array, are then clocked out slow-scan (20 ms/line at 50 kpixels/sec). The average signal level is then measured with a computer and plotted. As seen in the figure, well capacity is optimum for line transfer periods > 1 micro-sec. This occurs at $+V_V = 3$ V and is where BFW = SFW. For shorter periods full well degrades and the optimum point moves to a higher clock potential.

Figure 12. Well capacity as a function of clock voltage and line transfer time for the Cassini imaging CCD. When line rates become faster than 1-micro-sec the full well curve begins to change shape as demonstrated here. Clocked too fast causes charge to bloom backwards from transfer direction. Optimum full well occurs at a higher clock voltage at a reduced level. Charge packets less than full well can be transferred at the speeds indicated without difficulty.



Degradation of full well observed in Figure 12 is attributed to fringing fields between phases (refer to top illustration in Figure 13). Fringing fields play an important role in transferring charge. Without these fields charge only moves by thermal diffusion aided by self repulsion. Fringing fields are greatest when the potential wells are empty. As charge collects the potential difference between phases decreases resulting in smaller fringing fields. It is for this reason why it is easier for the CCD to transfer a smaller charge packet than a large one. As full well is approached fringing fields essentially disappear. Under full well conditions extra time must be given to allow charge to thermally diffuse from phase to phase. If the clocking rate is too fast charge will bloom backwards as illustrated in the lower illustration of Figure 13. Figure 14 presents three column traces for the CCD characterized in Figure 12 demonstrating that charge always blooms opposite to transfer direction when the CCD is clocked too fast. Charge moves from right to left in figure.

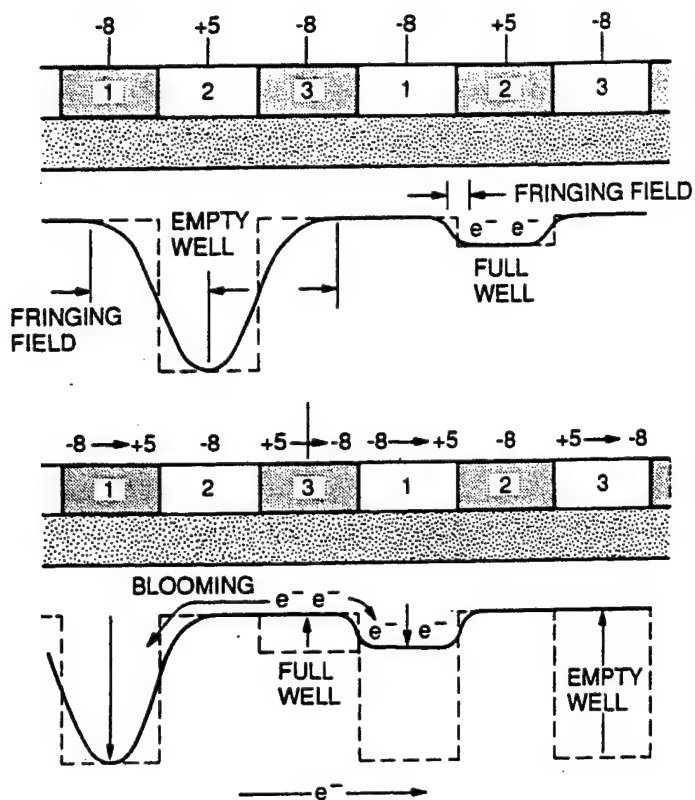
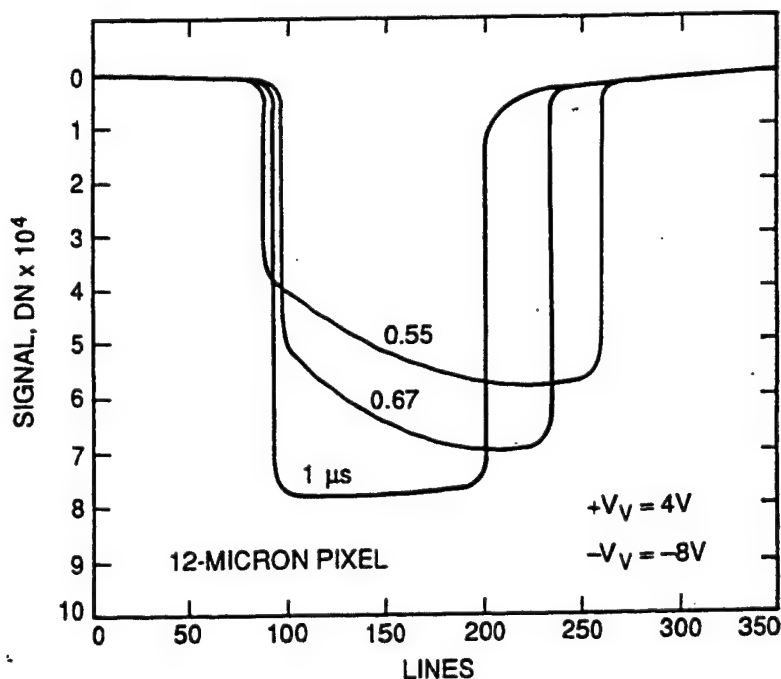


Figure 13. Top illustration shows relative fringing field strength for an empty and a full potential well. When a well is filled with charge the potential difference between the collecting phase and barrier phase is much smaller compared to when it is empty. This reduces the electric fields between them. Charge that remains in a collapsing phase that was full must primarily transfer by thermal diffusion and cannot rely on fringing field. Smaller charge packets receive the benefits of fringing fields, the reason why they can be transferred much faster. The bottom illustration shows why blooming occurs if the CCD is clocked too fast. Charge that remains in a collapsing phase blooms backwards.

Figure 14. Data taken from the same CCD as Figure 12 showing that blooming caused by high speed clocking is always backwards from transfer direction. Vertical clocks switch from -8 V (inverted) to +4 V.



Fringing fields can be increased by elevating the clock potential (i.e., V_V). Full well will increase initially until SFW is reached at which point it then decreases. This effect is observed in Figure 12. Note also full well characteristics are independent of clock rate when SFW limited.

Figure 15a is an x-ray response exhibiting perfect CTE for a line transfer rate of only 200 ns. The charge packets transferred are very small (1620 e-). As Figure 12 shows CTE should be well behaved when charge packets this size are transferred. PISCES modeling programs predict that CTE will become limited for line times of a few nano-seconds for the CCD tested. Unfortunately it is very difficult to generate three phase clocks this narrow to verify this prediction. For signals of this magnitude CTE is not limited by the CCD but by the clock drivers.

Figure 15b shows an x-ray response when the horizontal clocks are only switched from -0.5 V to +0.5 V. Fringing fields under these clocking conditions are very small. CTE is well behaved as long as the clocks are slow (less than 0.1 M pixel/sec).

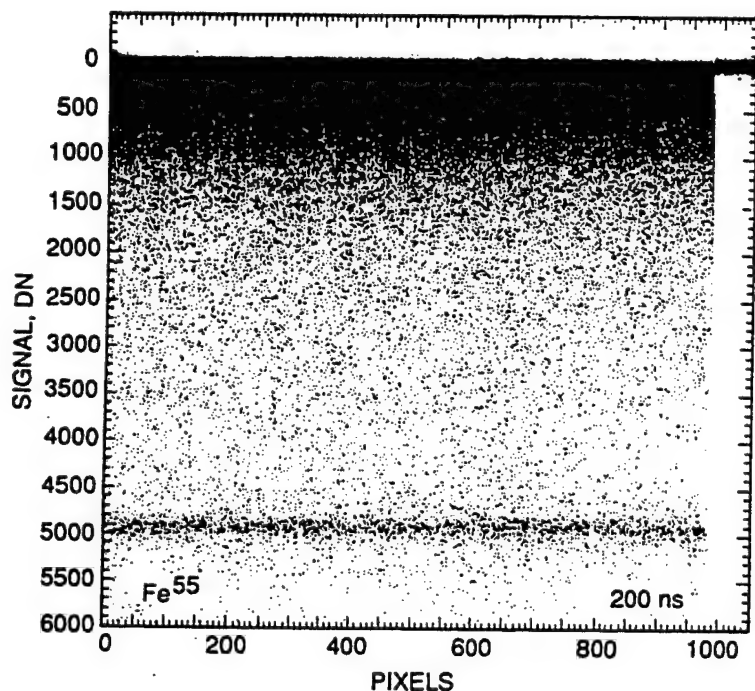
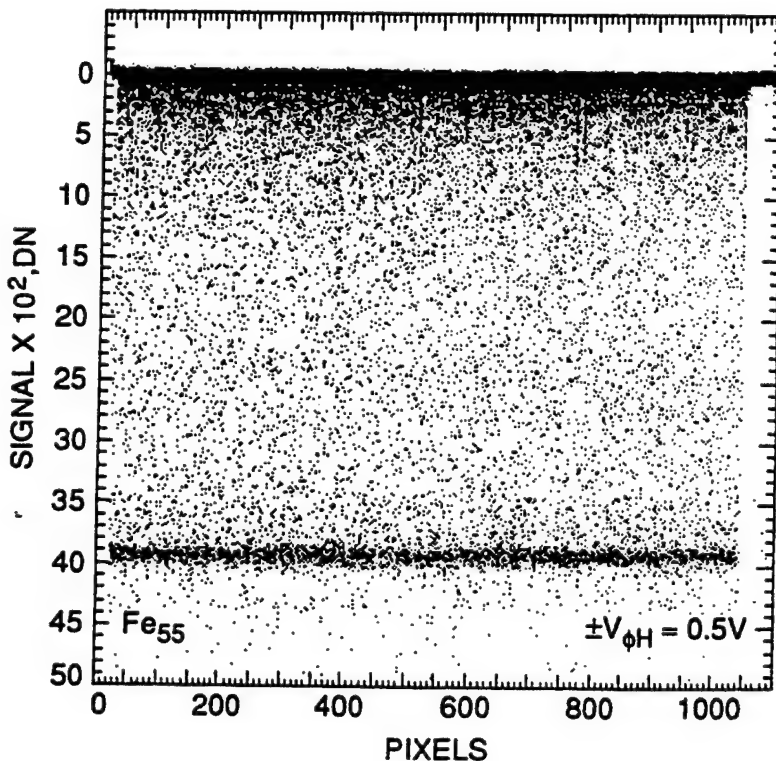


Figure 15a. HCTE Fe-55 x-ray response for the Cassini CCD. Ideal VCTE is demonstrated. Lines are transferred at 200 ns, much faster than the rates tested in Figure 12. In that the maximum charge level is only 1620 e- fringing fields are primarily responsible for transferring charge. Theory indicates that line times ten times faster than this are possible in transferring x-ray events this small.

Figure 15b. HCTE Fe-55 x-ray response for the Cassini CCD. In this test the horizontal clocks only swing 1 volt. This effectively reduces the fringing fields and transfer speed. Data was collected at 50 kpixels/sec using 8 micro-sec to make a pixel transfer. Rates faster than 0.1 M pixels/sec will result in degraded CTE performance. Increasing the clock swing amplitude allows faster transfer rates.



The data shown in Figure 12 was generated for a three-phase 12-micron pixel allocating 4 microns/phase. It is more difficult to transfer charge when pixel dimensions increase in size since fringing fields are less influential in the center of a phase (for an empty well a fringing field may extend a couple microns into an adjacent phase). Figure 16a plots well capacity for a 15-micron pixel CCD (1024 x 1024, 3-phase CCD). Note that full well begins to degrade for a longer line transfer period compared to a 12-micron pixel. Figure 16b shows full well characteristics for a 7.5-micron pixel (800 x 800, 3-phase). Charge transfer for this CCD is much faster. Full well reduction does not take place until the line time is shorter than 300 ns. Figures 12, 15 and 16 show that CCD speed is limited by pixel size and charge level transferred.

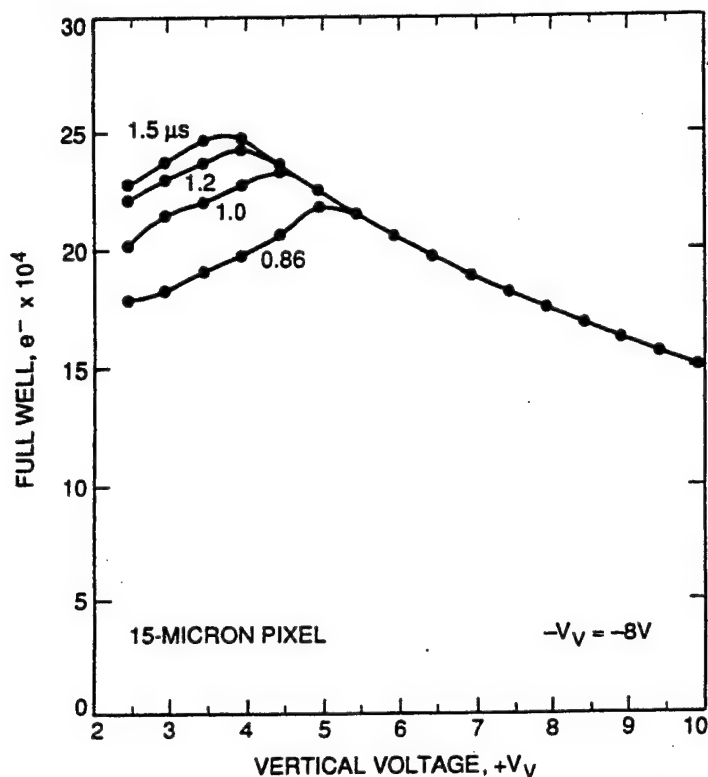
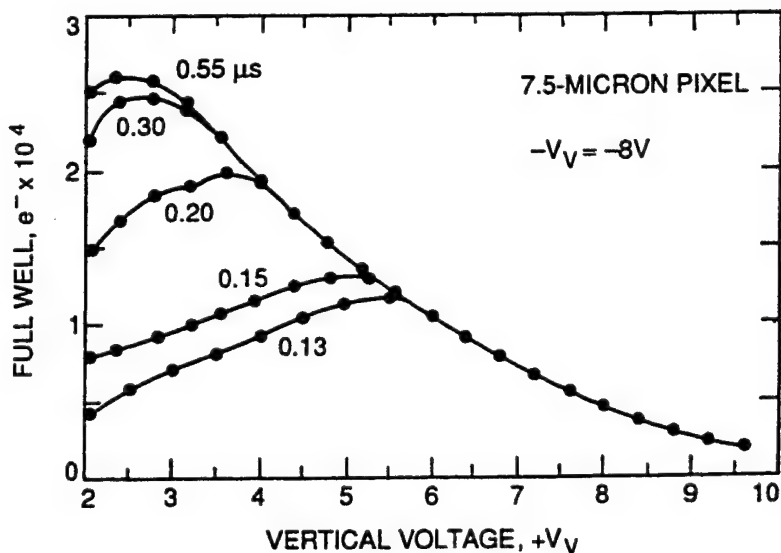


Figure 16a. High speed full well curves generated by a 800 x 800, 15-micron pixel WF/PC II CCD. Note that full well degrades for a longer line time compared to Figure 12 (generated for a 12-micron pixel device). Fringing fields only extend into a neighboring phase by a couple microns when the phase is empty. When the gate length is increased the fringing field strength at the center of the phase is reduced. Charge there must transfer by diffusion increasing line time requirements.

Figure 16b. High speed full well curves generated by a 7.5-micron pixel WF/PC II CCD. Note that the full well curve does not begin to change shape until line times shorter than 0.55 micro-sec are tested. The gate length for this CCD is small (2.5-microns) resulting in strong fringing fields at the center of a phase resulting in greater speed. High speed CCDs are typically small pixels devices unless multiple phases are used to shorten gate length.

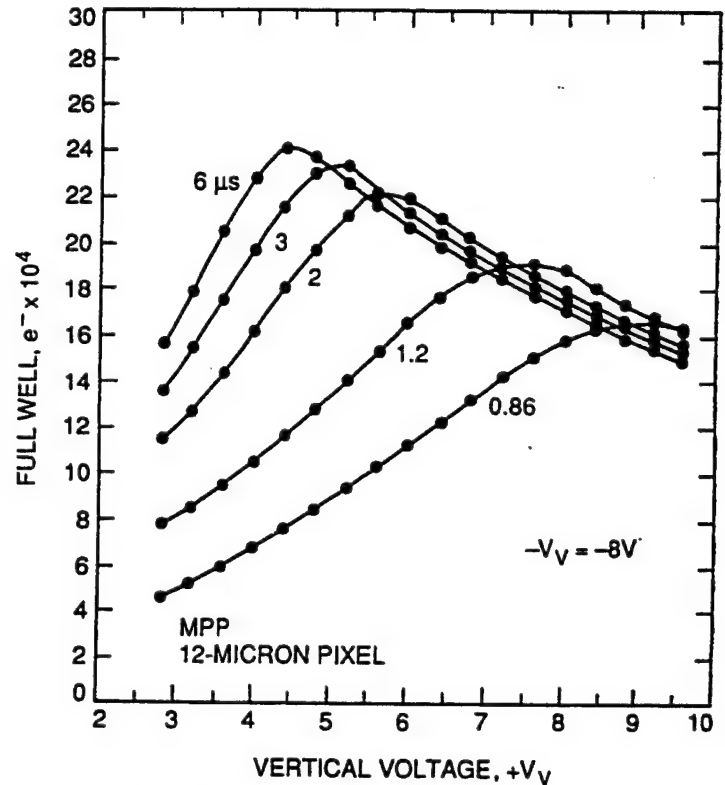


MPP CCDs clocked fast-scan typically exhibit CTE problems. The difficulty has been traced to the MPP implant itself. During high temperature processing the MPP implant encroaches under adjacent phases (i.e., phases 1 and 2). This effect decreases the potential at the edges of these phases generating a small but influential potential barrier to electrons. When phase 3 is empty

these barriers are overwhelmed by fringing fields and charge transfer is well behaved. However, as signal electrons are collected the fringing fields collapse and the barriers become apparent making it more difficult to transfer charge. Therefore, the full effect of this problem becomes present at full well conditions where fringing fields are smallest. Charge can only escape over the barriers by thermal diffusion. Since the potential associated with the barriers are probably many times greater than kT/q , significantly more transfer time is required compared to CCDs without MPP.

Figure 17 shows full well characteristics for a 12-micron pixel MPP Cassini CCD. Note that transfer speed is significantly reduced compared to the Cassini CCD characterized in Figure 12 without MPP (for the same pixel size). In fact, optimum full well is never achieved for the longest transfer times tested. For this reason alone MPP technology is not incorporated in the Sandbox CCDs because transfer times less than 1 micro-sec are required.

Figure 17. High speed full well curves generated by a Cassini MPP CCD (12-micron pixel). Note that the curves are considerably different than the full well curves shown in Figure 12 for a non MPP CCD. As the text explains, the MPP implant slows down transfer time. Line transfer periods > 10 micro-sec are required before full well characteristics stabilize. Vertical pixel summing in the horizontal register is used to readout the CCD yielding full well figures approximately three times shown in Figure 11 (data collected at 300 K).



The responses presented in Figures 12-17 were generated by relatively small CCD arrays (i.e., $< 1024 \times 1024$). When larger arrays are characterized another high speed effect comes into play. Using the same testing sequence as described above, Figure 18 plots full well characteristics for a Loral Fairchild 2048 x 2048, 15-micron pixel MPP CCD tested at two line times (30, and 6 micro-sec). Also two regions of the CCD are interrogated, the center and edges of the array (indicated by "C" and "E"). Note that full well characteristics for the center of the CCD are worse than the edges. This is because the narrow clocks applied do not make it to the center of the array properly. The poly resistance of the gate electrodes and associated capacitance reshapes the clock reducing its amplitude (specifically the positive level since the negative side is driven deep into inversion). Although one can increase the clock voltage to reoptimize the center region of the CCD the edges would be driven into SFW. Each column of the CCD therefore has a different full well signature and optimization curve.

Figure 19 is a horizontal line plot for the same CCD characterized in Figure 18. The device is stimulated with a slant bar target in the region covering 1850 (H) x 2048 (V) pixels and driving it at 10 micro-sec/line (again using the same testing sequence described above but using a slant bar target). Ideally a flat response should be observed at the full well level indicated. However, as we move towards the center of the device full well decreases because of the poly resistance effect. Curiously there is an "extra dip" in full well at the left side of the chip. In this location there is a high impedance gate short associated with a single pixel. This short effectively lowers that gate voltage at the site and surrounding regions (the CCD is driven from both sides of the array). At slow-scan rates this problem is not apparent but becomes obvious for line rates < 30 micro-sec/line.

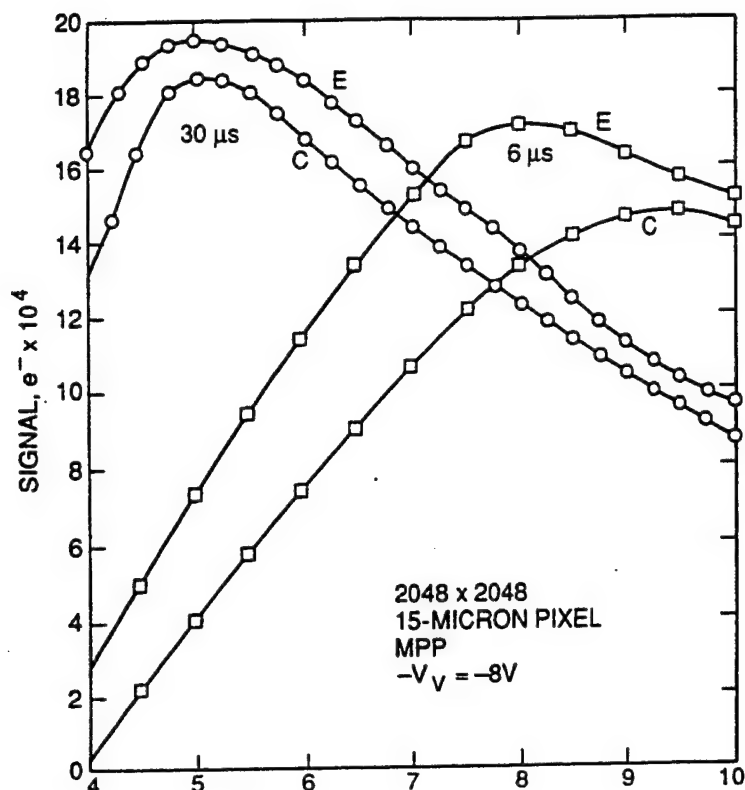
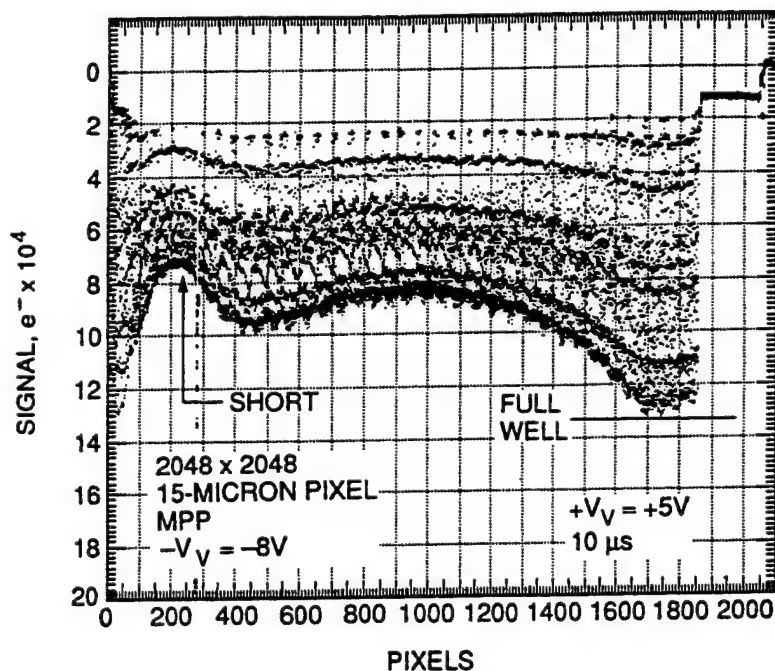


Figure 18. High speed full well curves generated by a 2048 x 2048, 15-micron pixel MPP CCD. Data is taken at the edge of the array (labeled "E") and the center of the array (labeled "C"). The edge is primarily limited in speed because of the MPP implant and the relatively large gate lengths employed (5-microns). Line transfer periods greater than 30 micro-sec are required to deliver optimum performance. The center of the array exhibits an additional problem caused by the long gate electrodes that feed the array. The poly resistance and capacitance associated with the gates results in a distributed R-C filtering effect that reduces clock amplitude and full well performance.

Figure 19. A slant-bar stacking response generated by the CCD tested in Figure 18. The vertical clocks swing from -8 V (inverted) to +5 V using a 10 micro-sec line time. These set of conditions limit full well performance over the array (refer to Figure 18). Note that full well decreases towards the center of the CCD because of the R-C clock filtering effect. In addition, a high impedance short is found on the left hand side of the array which perturbs the clock amplitude limiting speed in this location.



In summary, the CCD characterized in Figures 18 and 19 exhibits four high speed difficulties: (1) the fringing field problem; (2) the MPP barrier problem; (3) the poly resistance/capacitance problem and; (4) a high impedance gate short problem. The first difficulty results in blooming over the entire array and depends on pixel size. The second effect also causes similar blooming characteristics but is only associated with MPP devices. The third difficulty causes a shift in the full well curve towards a higher gate voltage when moving towards the center of the chip. The fourth difficulty is a localized problem and depends on the severity of the short.

The Sandbox CCDs were designed and will be processed to minimize these problems as much as possible. First, as mentioned above, MPP technology will not be employed. Second, either small pixels (e.g., 9-micron pixels for the Cinema CCD) or multiple phases (e.g., six phases for the Adapt II and Mach II CCDs) are used to increase fringing field strength. Third, aluminum bus lines are designed into the array to reduce the poly resistance/capacitance problem. For example, the Cinema CCD uses aluminum bus lines that are brought into the array every 256 columns and are strapped to the poly lines to reduce poly resistance. The fourth problem will be screened through testing.

5. QUANTUM EFFICIENCY

5.1 Introduction

To achieve the highest sensitivity possible, the Advanced Camera and Big CIT CCDs will be thinned and illuminated from the backside. This technique avoids the problem of photons being absorbed and lost in the frontside gate structure. Thinning also improves sub-pixel QE uniformity.

5.2 Advanced Camera CCD

Thinning will be performed by EG&G Reticon by standard techniques of chemical etching. Prior to thinning, the CCDs will be bonded to, and supported by, a glass substrate that covers and extends slightly beyond the pixel array. An unthinned border outside the pixel array will allow bonding of wires to the bond pads. The supporting substrate is necessary to strengthen the CCD to avoid damage during ground handling and launch, and to ensure flatness.

Thinning leaves a free silicon surface that oxidizes on the backside, leaving an accumulation of positive charge at the Si-SiO₂ interface. This charge creates a potential well at the back surface which traps photo-electrons, preventing them from reaching the frontside potential wells. This effect significantly reduces QE performance for photons that do not penetrate deeply into the silicon (i.e., short wavelength photons). Also the QE varies (i.e., hysteresis) when signal charge is collected in the backside well. This highly undesirable characteristic can be alleviated by a number of different passivation methods (e.g., backside charging, flash gate, biased flash gate, ion-implantation). For the Sandbox CCDs we will accumulate the backside using molecular beam epitaxy (MBE) delta-doping (Ref. 4).

The delta-doping method utilizes MBE to form a uniform and stable 1/3 monolayer of passivating boron atoms, on a 1-nm thick MBE grown silicon lattice both residing on the backside surface of the CCD. A thin layer of silicon is then grown on top of the boron layer. Advantages of the MBE approach are: (1) the boron layer can be precisely targeted to lie at a depth such that very strong electric fields are generated at the immediate surface that accelerates signal electrons to the frontside; (2) passivation over the area of the CCD is highly uniform; and (3) the boron atoms replace silicon atoms without distorting or damaging the silicon crystalline structure, so that annealing is unnecessary (required for ion-implantation). This technology achieves the "QE pinned" condition achieving 100 % internal QE. More-over the QE is uniform and stable over periods of many years (likely indefinitely).

An AR coating of SiO will be applied to the silicon surface to reduce reflection loss in the visible spectrum. The AR coating is expected to increase the overall QE of the device by a factor of about 1.4 in the visible wavelength range compared to noncoated CCDs.

High sensitivity for the Advanced Camera CCD must extend over a broad wavelength range (1100-11,000 Å). Unfortunately an efficient AR coating to cover this range has not been developed into the far UV. Therefore, a phosphor (lumogen) coating is used, the same coating that is used on the WF/PC II and Cassini CCDs. The phosphor works by absorbing short wavelength photons (<4800 Å) and re-emitting them at visible wavelengths (5400 Å for lumogen). The phosphor and AR coatings work together to enhance performance across the entire wavelength range and permits the CCD to be optimized for visible wavelength sensitivity. At visible wavelengths, where the phosphor is transparent, the AR coating enhances QE in the conventional way. In the UV and EUV, the phosphor absorbs photons and re-emits them at a wavelength where the AR coating is optimized (i.e., 5400 Å).

The photon re-emission for the phosphor is geometrically isotropic, so that half the re-emitted photons are directed away from the CCD. Thus, at best the QE in the region where photons are absorbed by the phosphor (1100-4800 Å) is one-half the QE where the photons are re-emitted (5400 Å).

Figure 20a presents an experimental QE plot generated by a backside illuminated WF/PC II CCD using the thinning and accumulation recipe described above. The CCD was thinned to approximately 8-microns which essentially thins the backside to the frontside depletion edge (the sensor is fabricated on 30-50 ohm-cm silicon). Note that the sensitivity below 4800 Å is essentially flat and is approximately half the QE achieved at 5400 Å. Also plotted in Figure 20a is a theoretical QE curve generated with an AR coating but without a lumogen layer. The parameters used best describe the WF/PC II CCD (7-microns of depleted material, 1-micron of field free material, QE-pinned, and 550 Å of SiO). Note that sensitivity for the experimental CCD is lower than theory between 3000-6500 Å primarily due to the lumogen layer and associated isotropic problem discussed above. However, below 3000 Å sensitivity is greater because of reflective and absorption loss related to the AR coating. Lumogen remains high (down to approximately a wavelength of 500 Å at which point the phosphor becomes transparent).

Also note that theory and experiment differ in the near IR. There are two possible explanations for this observation. First, the device may be slightly thicker than 7-microns resulting in slightly higher QE longward of 7000 Å. Second, the QE model employed (Ref. 5) does not include a reflection component at the frontside surface. That is, when near IR photons pass through the CCD some photons are reflected back from the front and make a second pass through the membrane again. This effect would increase sensitivity. In fact multiple passes result in fringing for the device always seen in the near IR for rear-illuminated CCDs.

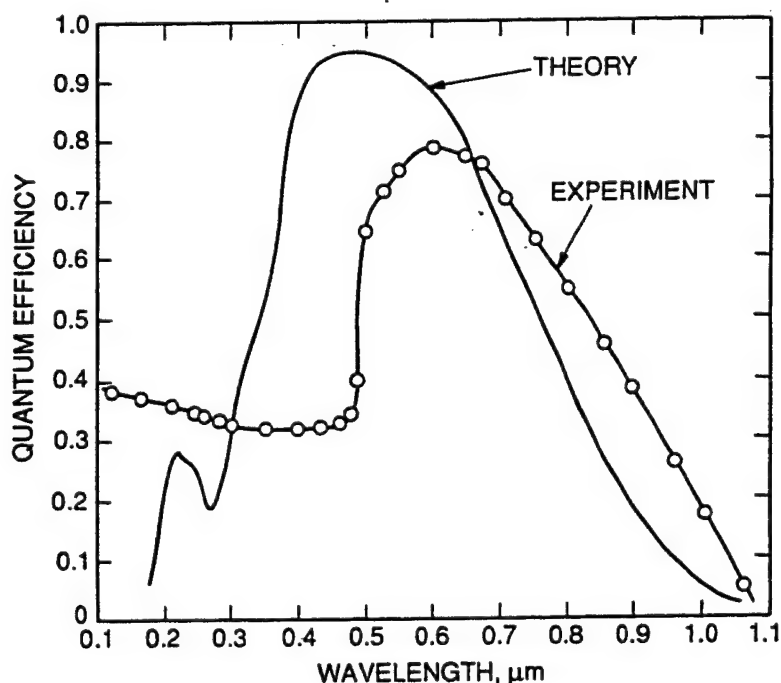


Figure 20a. QE curve generated by an experimental backside illuminated, MBE accumulated, SiO AR coated, lumogen phosphor coated, WF/PC II CCD. The AR coating thickness was adjusted to yield the highest QE possible at 5400 Å, the wavelength where lumogen fluoresces. The sensitivity in the UV (< 4600 Å) is approximately half of the QE achieved at 5400 Å. The theoretical QE curve shown uses parameters that best describe the CCD without a phosphor coating. Note the sensitivity is very poor shortward of 3000 Å where the AR coating becomes opaque. For this reason a phosphor coating is used on some Sandbox CCDs for extended response into the far UV. Lumogen responds to 500 Å where at this wavelength becomes transparent.

It is important that the Advanced Camera be thinned and backside illuminated because of the 9-micron pixel size employed. Figure 20b shows frontside QE responses for the Cassini and WF/PC II CCDs in the far UV (both CCDs are coated with 6000 Å of lumogen). The response for the Cassini CCD drops at 1800 Å where the quartz window above the CCD becomes opaque. Similarly the WF/PC II response degrades below 1500 Å where a MgF₂ window becomes absorbing. Note the QE for the Cassini CCD is lower than the WF/PC II device. The difference is because the WF/PC II pixel is 15-microns whereas the Cassini CCD is 12-microns. As the pixel dimensions shrink in size the gate overlaps become a greater portion of the pixel decreasing sensitivity. A gate overlap of 1-micron is used for all CCDs. Projected frontside UV QE for a 9-micron pixel is only 6 % considerably lower than the WF/PC II CCD (14 %). Hence, thinning is required if the Advanced Camera CCD is to outperform the WF/PC II CCD. Figure 20a shows that QE will be in excess of 30 %.

5.3 CIT CCD

The Big CIT will be thinned and processed in the same fashion as the Advanced Camera CCD except that the lumogen layer will not be deposited. The CCD will be used in ground-based astronomical applications, and therefore, UV sensitivity only needs to be high to 3000 Å (the wavelength where the atmosphere cuts off). The expected QE for the CCD will follow close to the theoretical curve shown in Figure 20a (the AR coating will be slightly thinner to boast UV QE performance with some sacrifice in the visible).

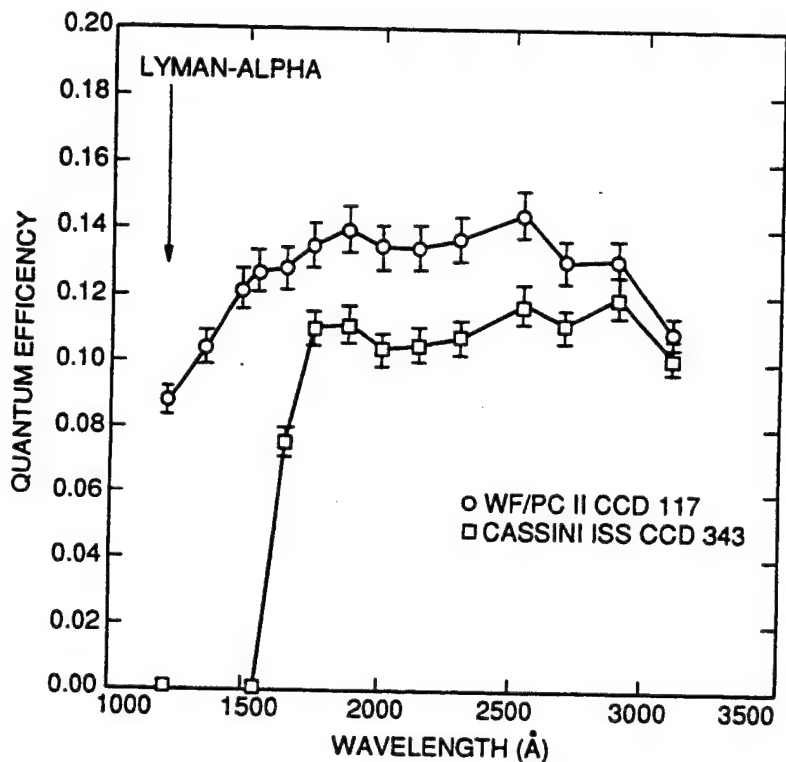


Figure 20b. Experimental QE plots generated by a WF/PC II and Cassini frontside illuminated lumogen coated CCDs. The WF/PC II response is higher because the pixel is larger than the Cassini CCD (15-microns compared to 12-microns). Gate overlap (approximately 1-micron) between phases becomes a greater percentage of pixel area for smaller pixel devices reducing QE at 5400 Å. For small pixel (<12-microns) devices that are coated with a phosphor this effect becomes important requiring thinning and backside illumination.

6. READ NOISE

Curiously, read noise for the CCD is still limited to more than 1 e⁻ rms for slow-scan operation. Some CCDs exhibit noise floors as low as 1.5 e⁻ slightly better than the first lightly doped drain (LDD) CCDs fabricated several years ago. Flicker (1/f) noise generated by the on-chip amplifier limits read noise. Many theories have been presented to explain where 1/f noise is actually generated in the output MOSFET. One popular theory is that carriers that make up the drain current tunnel between the buried channel and surface states primarily near the source of the MOSFET. Presumably, the trapping and detrapping of charge at the interface is the noise voltage associated with 1/f noise. Although this theory may be correct a solution to eliminate 1/f noise is not forthcoming. Nevertheless, 1/f noise has been reduced compared to MOSFETs fabricated in the past by improved surface passivation techniques (the corner frequency for 1/f noise is approximately 50 kHz at room temperature). It now appears that a 1/f limit has been reached, and therefore, we must live with 1/f noise and remain slightly above 1 e⁻.

Floating gate technology has allowed the CCD to circumvent the 1/f noise problem by reading pixels nondestructively (amplifiers of this type are called Skipper amplifiers). Multiple samples of a pixel are averaged together to improve noise performance (noise is reduced by the square-root of the number of samples taken). The technique is very effective and has allowed noise levels of 0.3 e⁻ rms to be achieved thus breaking the 1 e⁻ barrier. For best results the CCD video signal is processed using the technique of correlated double sampling (CDS). An optimum sample-to-sample time is found for best S/N depending on spectral noise characteristics of the output amplifier (i.e., white and 1/f). We generally find for today's CCDs that when sample times > 4-micro sec that correlation 1/f noise frequencies begins to be lost. At this point read noise does not decrease by the square-root of sample-to-sample time because of 1/f noise skirt. Therefore, multiple samples are taken using

a sample period of 4-micro sec.

The general trend for reducing read noise at the manufacturer has been to increase the sensitivity of the output amplifier (i.e., e/V) without significantly increasing white and $1/f$ noise. This has been accomplished by primarily reducing gate capacitance of the output amplifier. This is accomplished by reducing its size and using LDD technology. Although $1/f$ and white noise has increased slightly in the optimization process the net noise (in rms e^-) has been reduced due to the sensitivity increase. Sensitivity is approximately 4 micro- V/e^- for 2-3 e^- amplifiers, about a factor of 10 times higher compared to the WF/PC I CCD which exhibited a read noise of 13 e^- rms. Increasing the sensitivity any more than this results in a significant $1/f$ and white noise increase and a corresponding increase in read noise.

As mentioned above, read noise for high performance floating diffusion CCD amplifiers is about 2 to 3 e^- rms. Skipper amplifiers typically require between four to nine samples/pixel before a 1 e^- -noise floor is achieved. Two Sandbox CCDs (Adaptive Optics and Advanced Camera) will utilize Skipper amplifiers. The current WF/PC II CCD is achieving 4-5 e^- (about 3 times better than the WF/PC I CCD). The Advanced Camera CCD employs two floating diffusion amplifiers for single sampling that should yield 2-3 e^- improving noise by a factor of 1.5 or more. The device also employs two Skipper amplifiers for improved noise performance if desired.

The sensitivity improvement has resulted in a side problem related to linearity. As the sensitivity increases the voltage swing on the gate of the output MOSFET increases. This swing can be quite large if the CCD is used over its full dynamic range. For example, the Cassini CCD exhibits a sensitivity approximately 4 micro- V/e^- and a well capacity of 134,000 e^- thus swinging the gate by -0.54 V. This voltage variation results in a nonlinearity because the gate bias point for the amplifier shifts slightly to a new gain level. The Cassini CCD is limited to approximately 130,000 e^- before a nonlinearity of 1 % sets in. The on-chip amplifier therefore exhibits a dynamic range of its own bounded by a linearity requirement (the dynamic range for the Cassini output amplifier is approximately 26,000 assuming a read noise of 5 e^-). Nonlinearity is not a problem for the 9-micron pixel Sandbox CCDs because full well is less than the Cassini CCD (approximately 60,000 e^-). The amplifiers are nearly identical in size and input capacitance.

The Cinema CCD uses a three-stage output amplifier for high speed operation (Figure 1a). The first stage MOSFET is extremely small in size and will exhibit high sensitivity (> 8 micro- V/e^-). Because of its size the amplifier will exhibit a high $1/f$ noise cut-off frequency. However, the horizontal register will be read out at approximately 8 Mpixels/sec requiring a sample-to-sample time < 50 ns. Therefore, $1/f$ noise will be correlated and removed by CDS. The second and third stages are used to increase bandwidth and drive characteristics.

7. IONIZING RADIATION DAMAGE

7.1 Introduction

This section briefly discusses some new concerns in reference to ionizing radiation damage for the CCD. Much could be said about this field of study because it is still unfolding. Discussions given in this section are important to the Advanced Camera and Pluto Flyby CCDs because they will be used in space where high energy radiation sources are found.

7.2 Full Well

Figure 21 shows a full well plot generated by a Cassini MPP CCD before and after being irradiated with 6 krad of Co-60 gamma-rays. A significant full well loss is observed considering the device was exposed to such a small amount of radiation. Also note that the optimum full well point shifts towards a higher gate voltage (2.2 V to 3.8 V), opposite to the expected direction. Flat-band shift is always negative for CCDs because radiation induces positive charge in the gate oxide. For example, a flat-band shift of -0.6 V was measured for this CCD. These peculiar results are now explained.

Figure 22 shows a cross-section view for the Cassini MPP CCD looking up the signal channel. The region where the channel stop and signal channel converge is referred to as the "birds beak" region. The field oxide over the channel stop region is approximately 20 times thicker than the gate oxide. During irradiation the field oxide absorbs more radiation than the gate oxide. In turn the flat-band shift in the field oxide region is greater than compared to the middle of the channel.

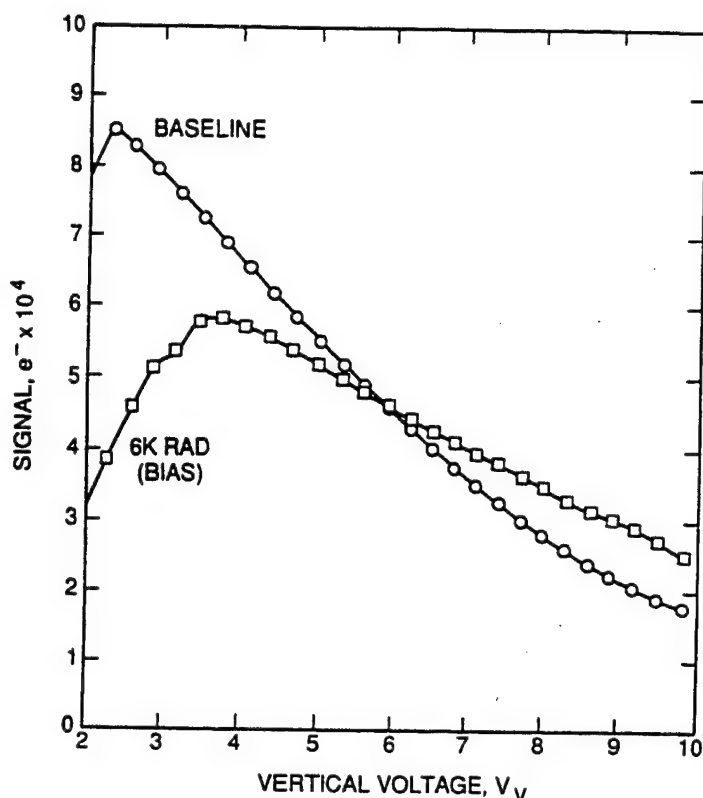


Figure 21. Full well characteristics for a MPP Cassini CCD before and after being irradiated with 6 krad of 2 MeV gamma-rays. Full well degrades because of a significant flat-band shift that develops near the channel stop region as discussed in the text. The device was biased during irradiation which intensifies the problem.

Figure 22. Cross-sectional view of signal channel and "birds beak" region. Field oxide above the channel stop captures more incident radiation resulting in more damage in the region. As a result, the flat-band shift is greater causing the potential to change more relative to the center of the channel. This effect significantly influences BFW and inversion characteristics for the CCD.

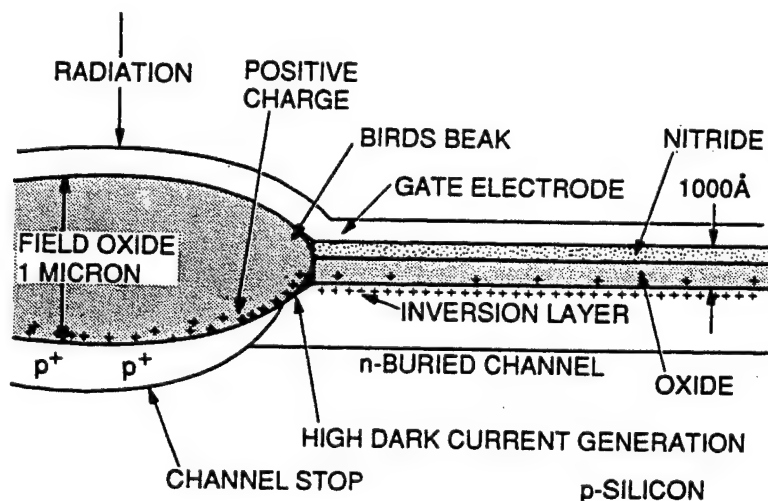


Figure 23 demonstrates the magnitude of shift that occurs in the birds beak region after the CCD is exposed to a small amount of radiation. Figure 23a plots dark current as a function of $-V_V$ when the gates to the upper half of the array are grounded during irradiation. Figure 23b is a similar plot except that the lower half of the array is biased at -9 V (the CCD is designed as a frame transfer device so split bias during irradiation is possible). A baseline response is shown in both figures before irradiation. These responses show that the inversion point occurs approximately at $V_V = -6.0$ V. After irradiation the inversion point shifts due to positive charge build-up in the oxide. The shift is significant for the biased portion of the CCD (for reasons explained below). The shift in both cases is much greater than the flat-band shift related to the channel itself (-0.6 V).

Knowing these effects we return to the full well issue presented in Figure 21. Figure 24 presents PISCES potential diagrams for the Cassini MPP CCD. The top diagram plots the channel potential for a non MPP phase beginning in the center of the channel (labeled $x=2.6$ microns) and moving towards the birds beak region ($x=0$). The analysis assumes a flat-band voltage of zero volts (i.e., pre-irradiation) and a gate potential of -8 V. Note the channel potential is maximum in the center and collapses

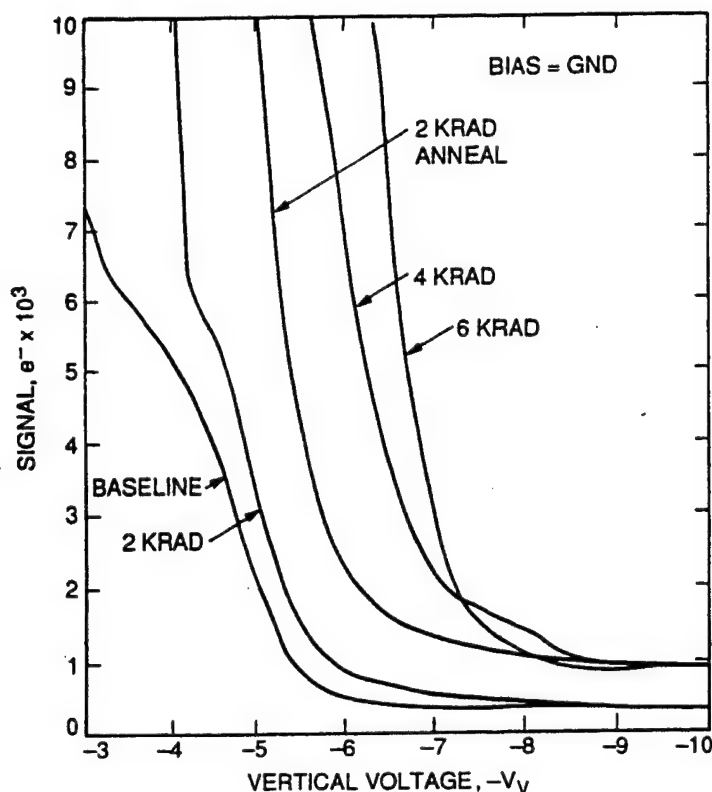
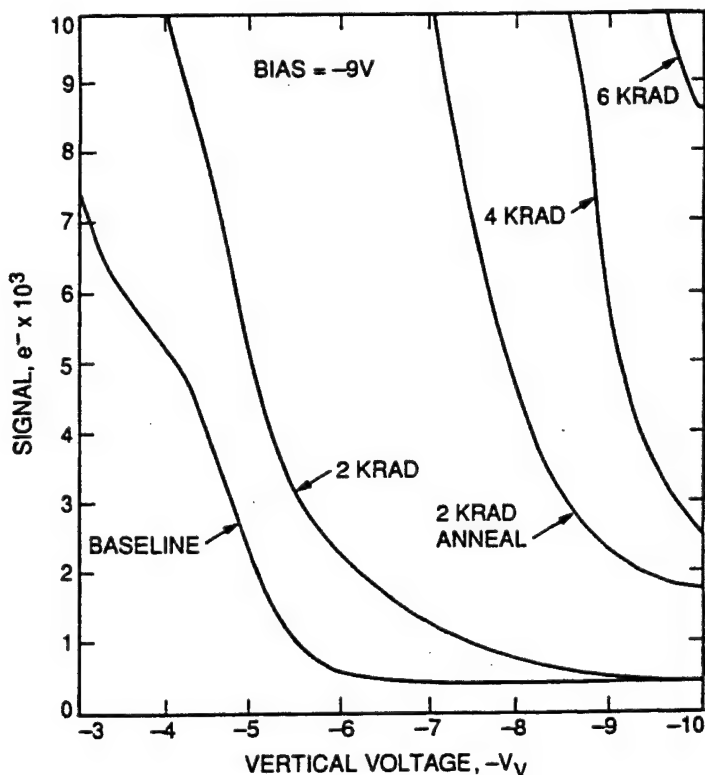


Figure 23a. Dark current plots as a function of negative clock drive for three Co-60 irradiation levels. Note that onset of inversion requires additional drive for increasing amounts of radiation. At 6 krad the shift is 2.5 V, greater than the a 0.2 V shift measured in the center channel. A short high temperature anneal is performed after 2 krad which increases dark current for reasons explained in text (i.e., reverse annealing).

Figure 23b. The same data plots as Figure 23a except the gates are biased at -9 V during irradiation as opposed to ground potential. Bias to the CCD separates electron-hole pairs generated in the gate insulator during irradiation resulting in less pair recombination and more damage. The resultant damage is considerably more than the unbiased case shown in Figure 23a. The 6 krad curve shows that the flat-band shift is significant possibly as high as 15 V in the region. Flat-band shift in the center of the channel is only 0.6 V.



to zero at the channel stop region. The lower diagram plots potential for the MPP phase with oxide charge added thereby simulating radiation induced charge. Note that the potential now increases as we approach the birds beak because the pinned condition cannot be maintained due to positive charge build-up (gate potentials as high as -24 V are required to pin the region). As discussed above, full well is determined by the potential difference between the MPP phase (phase 3) and non MPP phases

(phases 1 and 2). As indicated in the figures after irradiation only a 0.7 V barrier is left (compared to a 2 V difference). Hence, BFW is reduced considerably near the birds beak compared to the center of the channel (refer back to Figure 21). Since BFW is reduced the optimum full well point shifts towards SFW.

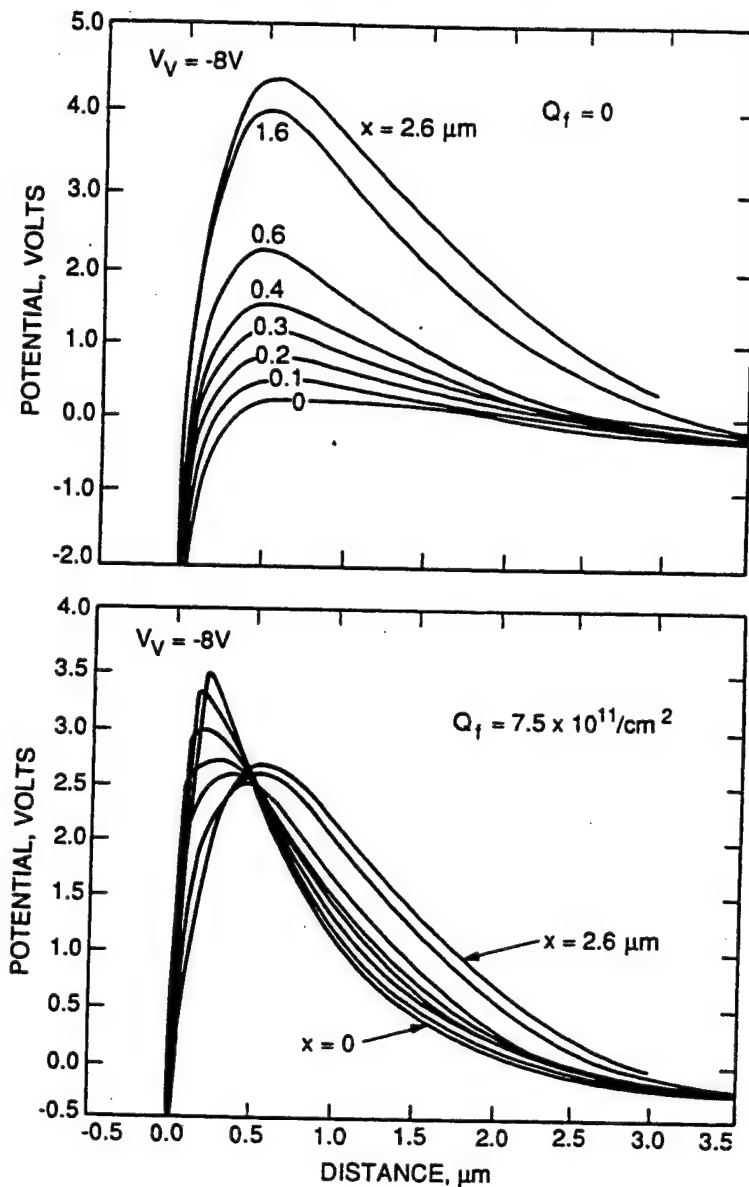


Figure 24. PISCES modeling data showing potential plots for the signal channel and the birds beak region. The top potential plots are for a non MPP phase without oxide charge. The potential at the center of the channel ($x=2.6$) is 4.3 V and the channel stop region ($x=0$) is 0 V. The lower plot shows the potential distribution after oxide charge, Q_f , is added simulating radiation induced charge. Note that the potential in the birds beak region is now much greater because of positive charge build-up in the region (the region becomes unpinned for the gate voltage applied). Flat-band shift in the center of the channel is small and negligible. The potential difference between phase 3 and phases 1 and 2 is therefore reduced considerably. Premature blooming occurs along the edge of the channel stop region reducing overall full well performance (Figure 21).

The inversion and full well problems caused by radiation can be reduced by employing a super notch structure as shown in Figure 25 (the "notch" structure indicated is employed to reduce bulk trapping - an effective technology). Super notch effectively isolates the birds beak region from the signal channel. For example, Figure 26 plots dark current as a function of negative drive for different super notch widths ranging from 2-15 microns (pixel pitch is 18-microns based on a 4-micron channel stop before processing). The test CCD was only slightly damaged with radiation to demonstrate flat-band shift at the birds beak region. Note when the channel butts up to the channel stop (i.e., 15-micron super notch) that it is difficult to invert the CCD requiring a negative drive voltage of -10 V. However, when the channel is pushed back by 4-microns from the center of the channel stop inversion is controlled (requiring -6.5 V).

The penalty paid in implementing super notch is a reduction in full well because the width of the signal channel is reduced. Figure 27 plots well capacity for the same CCD. Full well is reduced from 270,000 e⁻ to 195,000 e⁻ when a 10-micron super notch is employed. The Sandbox CCDs will not employ super notch because of full well reasons. However, the studies above

(and below) have guided us around the birds beak flat-band shift problem by the manner in which the CCD is clocked as discussed in the next section.

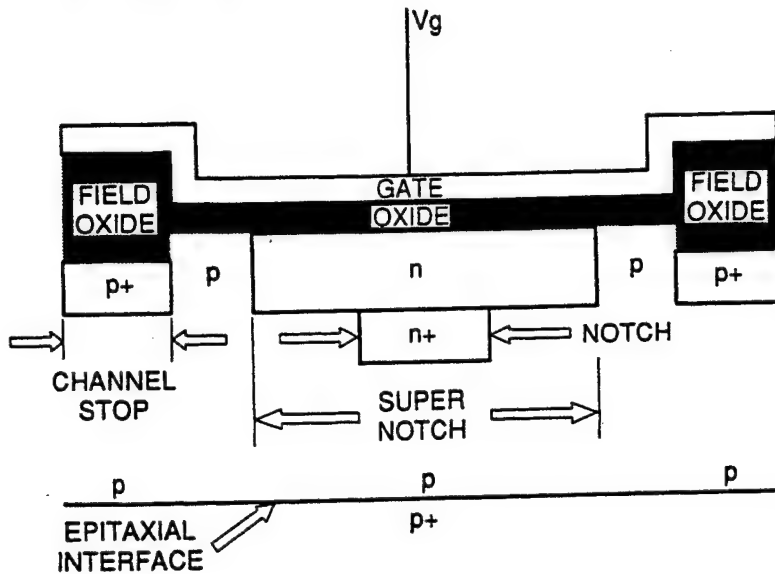
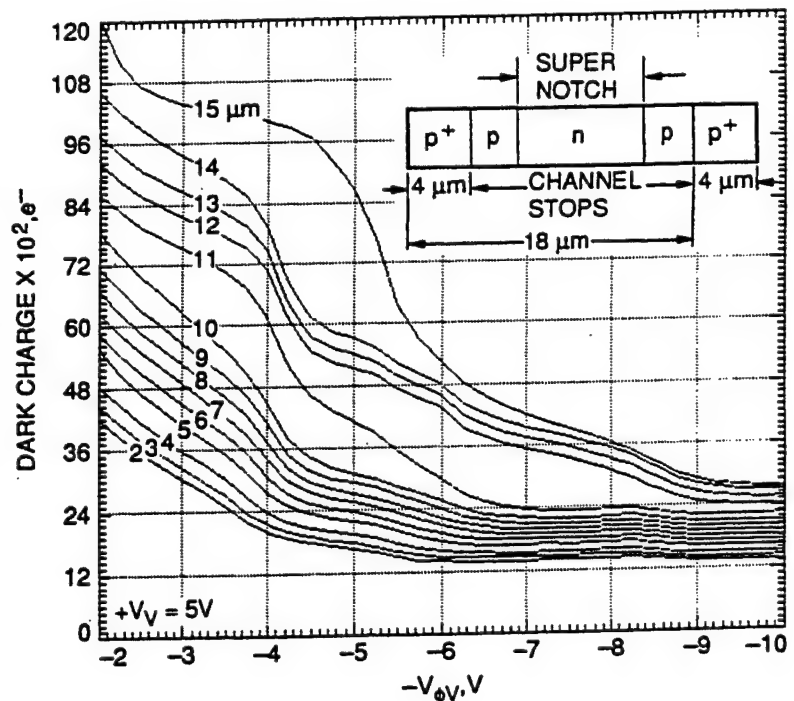


Figure 25. Super notch structure used to isolate birds beak and signal channel. The technology reduces flat-band and inversion problems induced by radiation. The buried channel is defined by a separate mask leaving sufficient p material between the p⁺-channel stop and n-buried channel regions. The "notch" structure also shown is another mask used to implant the signal channel deeper and provide a smaller channel to transfer small charge packets. The notch is often used to reduce CTE trapping problems for ultra-large CCD arrays or devices used in radiation environments.

Figure 26. Dark current plots demonstrating the benefits of super notch after the device is exposed to a small amount of ionizing radiation. Various super notch channels are investigated on this test CCD that gradually show the isolation effect as the width is varied. A 15-micron super notch overlaps the channel stop and provides no isolation. In this case it is difficult to control inversion requiring a clock drive > -9 V. However, as the super notch width is made smaller isolation is achieved. A width of 10-microns provides good isolation, approximately 1-micron removed from the channel stop after processing.



7.3 Reverse Annealing

Dark current results for the 4 krad and 6 krad curves presented in Figure 23 are actually much worse than shown because of an interesting effect referred to as reverse annealing. Note that a significant shift occurs for the 2 krad curve after the CCD was annealed at +50 C for 2 hours. Annealing was not performed after the 4k and 6 krad doses but given time they would shift significantly also.

Reverse annealing is a radiation effect that has been recently studied in detail for the Cassini and Space Telescope projects. It appears to be a very complex, long term chemical reaction associated that takes place in the gate insulator immediately after

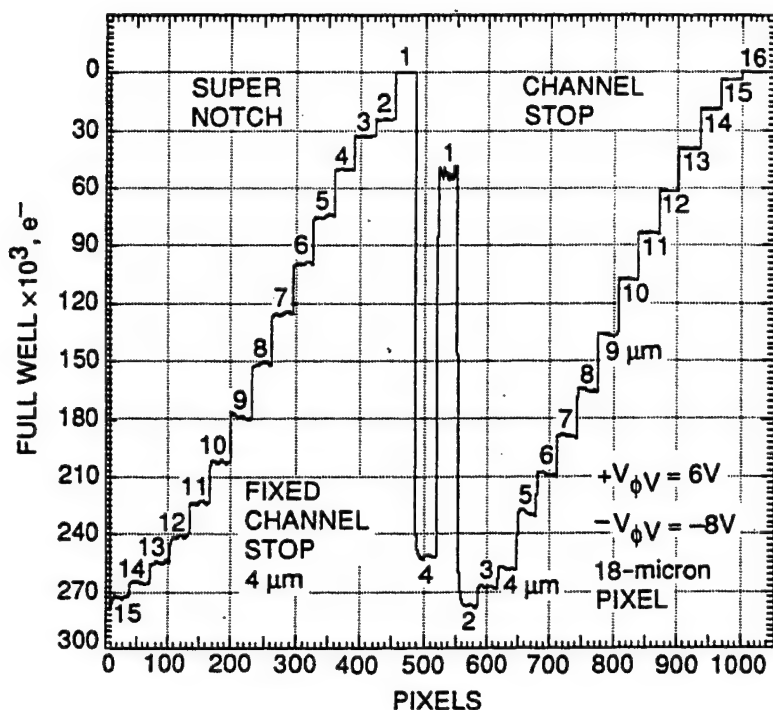


Figure 27. Full well line trace generated by the same test device as Figure 26 designed with different super notch widths in columns 0-500. Note as the width of the notch decreases a corresponding decrease in full well is experienced. Good channel stop isolation requires a 10-micron super notch. Full well is reduced from 270,000 e- to 195,000 e- when this super notch is employed. Super notch is not used in the 500-1000 column region. In this area the channel stop width is varied from 1-micron to 16-microns and is used to find the minimum channel stop width. For the CCD process used 2-micron channel stops are sufficient.

irradiation. Similar effects have been observed in other types of MOS devices and various models presented (Ref. 6). We briefly present one of these models here that appears to fit our observations.

The process begins when radiation induces electron-hole pairs in the gate insulator (wavelengths shortward of 1800 Å where oxide absorbs incoming photons). Most of these carriers recombine, however, some holes remain and slowly diffuse in the insulator. Through several chemical steps the free holes generate H^+ ions in the layer. These ions diffuse towards the Si-SiO₂ interface and attack it with the following reaction taking place:



where 3H-Si-H is a hydrogen passivated trap (after device fabrication the CCD is passivated with hydrogen gas at a high temperature to reduce interface state density). The free electron in the equation comes from the silicon substrate via tunneling. Note the end result of this process is a dangling bond or interface state.

The hydrogen ion diffusion process takes time and is exponentially dependent on operating temperature. For example, we project that 4k and 6krad dark current curves shown in Figure 23 would require several years to stabilize at room temperature. However, warming the CCD to +50°C reduces this time to a couple of hours. Once formed, the new states increase dark current generation.

It is not too surprising that the amount of reverse annealing measured is also dependent on bias conditions to the CCD because holes and ions are both charged. For example, when a negative bias (relative to substrate) is applied an electric field is generated that attracts the H^+ ions to the gate instead of to the interface stopping the reverse annealing effect. However, biasing the CCD positively repels the ions to the interface accelerating the reaction. Figure 28 shows how reverse annealing is controlled by the bias applied to the CCD during annealing after the sensor is irradiated. In this experiment the CCD is first irradiated with 2 krad of 5.9 keV x-rays. The CCD is then heated to 60°C and the dark current monitored (top plot). A region that is shielded from the x-rays and not damaged (bottom curve) is also measured to correct for temperature differences that occur in experiment as the CCD is thermally cycled. Note when the gates of the CCD are grounded (0 V) that dark current increases as positive ions make it to the interface. Biasing the CCD to +15 V increases the dark rate. However, applying -5 V stops the process completely as theory would predict. Returning the bias to +15 V causes the dark current to increase once again. The effect can

be completely controlled by the user.

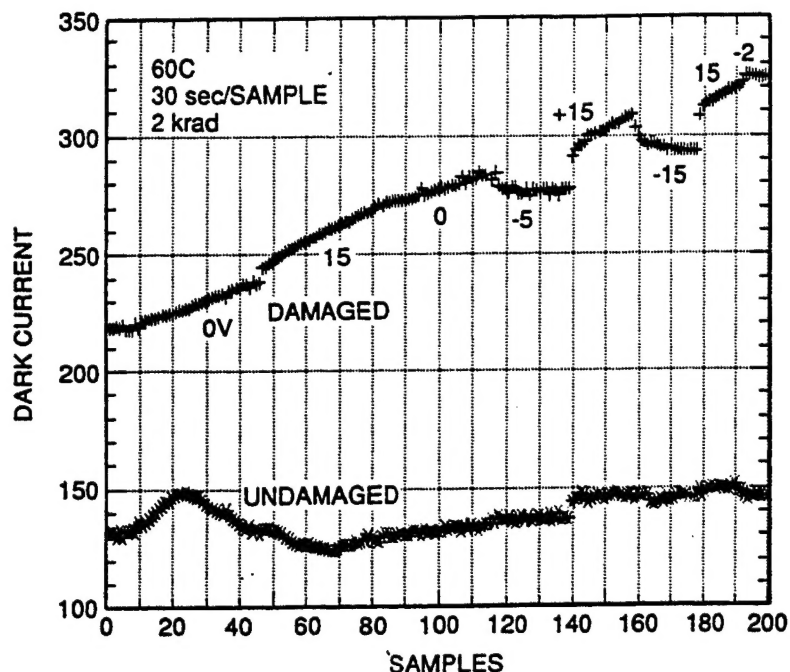


Figure 28. Dark current buildup as a function of time during a high temperature anneal cycle demonstrating the reverse annealing phenomenon. The CCD is irradiated and then annealed at 60 C as dark current increase is monitored. Note that the dark rate can be controlled by how the CCD is biased. The process can be turned off by biasing the CCD with a negative voltage (attracting positive ions to the gate). The lower plot is used to monitor operating temperature by measuring dark current in an undamaged region of the CCD.

We should also mention that flat-band also exhibits reverse annealing behavior. This is because the new interface states are electrically active and contribute to flat-band shift. Figure 29 plots SFW characteristics for a non MPP CCD during and after the sensor is irradiated with 5.9 keV x-rays. The baseline response before irradiation is indicated. Every 13 hours, while the device is being constantly irradiated with x-rays (with +15 V bias), a new full well plot is taken at room temperature as shown. Note that the full well curve slowly shifts towards the left indicating that new positive states are being created. After exposing the CCD to 46.5 krad it is annealed at +50 C for 2 hours and a new full well curve taken (triangles). A net flat-band shift of approximately 2 V is measured, half which comes from the high temperature anneal cycle.

Understanding the physical effects behind reverse annealing, and many other ionizing processes that occur in the CCD, one can minimize the amount of radiation damage induced. The improvement can be significant if carefully studied and planned (a factor of 20 was gained for the Cassini project simply by understanding the physics involved). For example, Figure 23 shows a dramatic difference in flat-band shift induced by how the CCD is biased. When biased during irradiation e-h pairs are separated leaving more holes to damage the interface. The unbiased condition allows carriers to recombine. Figure 30 shows the dark current increase for a Cassini CCD as a function of bias during irradiation. Minimum dark current generation occurs for zero bias because more recombination occurs. Note that damage is independent of polarity and is less for colder operating temperatures (carrier diffusion is less which promotes recombination). After irradiation it is important to keep the positive ions away from the Si-SiO₂ interface. This is accomplished by biasing the CCD negatively (normal clocking provides this condition automatically).

The CCDs characterized above were either damaged with gamma-rays or x-rays to a specified rad level (a rad is equal to 100 ergs of energy deposited per gram of material). Exposing the sensor to a different ionization source, such as protons, would give an entirely different result (for the same ionizing rad dose). The stopping power exhibited by protons is typically much greater than energetic photons. Hence, the density of e-h pairs generated by protons in the gate insulator is greater, and therefore, more recombination occurs. Hence, ionization damage for protons is significantly less than gamma-rays, a fact verified by experiment. This knowledge effectively reduces the rad level.

One concluding remark. The reverse annealing effect discussed above may not be observed for all CCD families. Ionization damage is very dependent on the CCD process employed (unlike bulk damage effects which for the most part are process independent). For example, it appears that reverse annealing naturally occurs in certain devices with Si₃N₄ (or similar) encapsulation layers. Devices that use an all oxide insulator, for example, may not experience the problem.

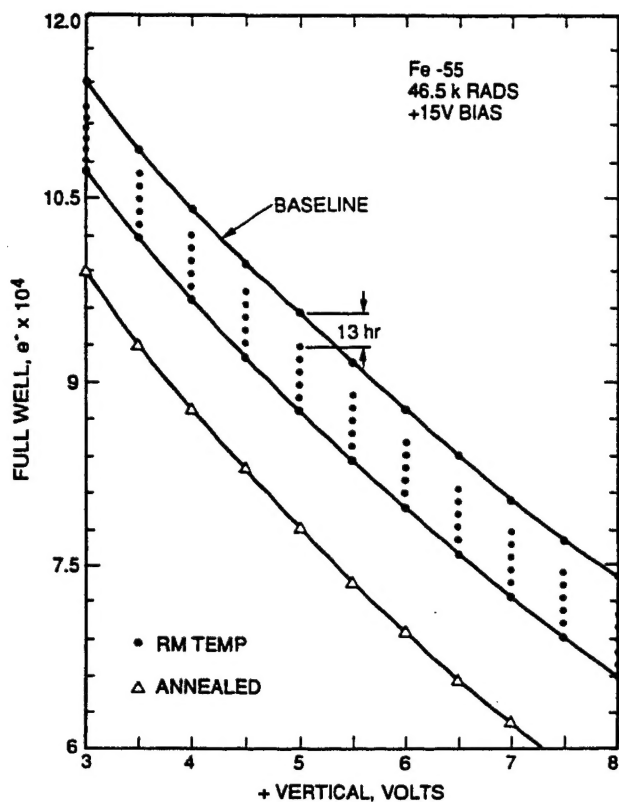
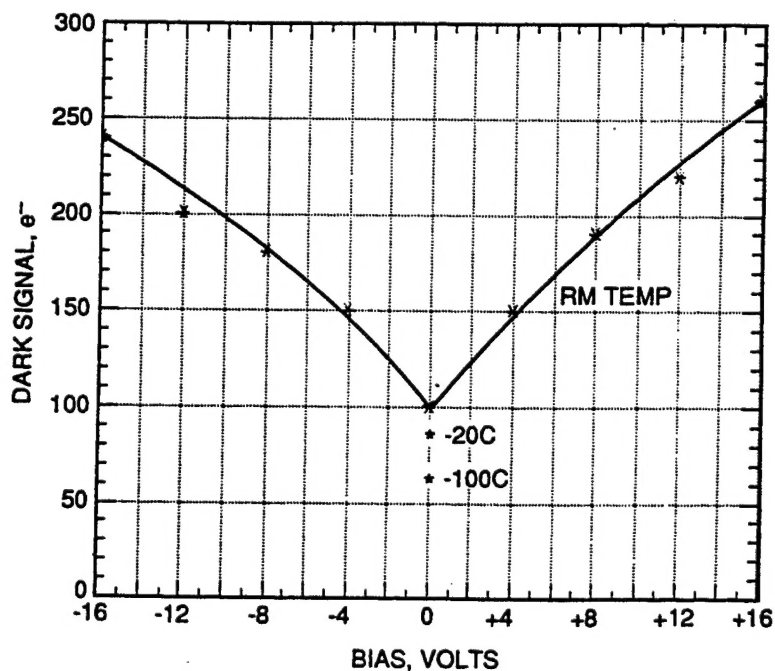


Figure 29. Full well characteristics before and after a CCD is irradiated with Fe-55 x-rays demonstrating that flat-band is also associated with reverse annealing. The plot labeled with triangles was generated after the CCD was annealed. As shown, a large shift in the full well curve was observed. The curve only plots SFW characteristics to avoid blooming problems that develop during irradiation.

Figure 30. Dark current as a function of CCD bias during irradiation. Carrier recombination is minimized in the gate insulator when bias is applied by forcing radiation induced electron-hole pairs apart. Note that damage is independent of polarity as long as an electric field across the insulator exists. Irradiation polarity should not be confused with annealing polarity where different mechanisms are at work. For reverse annealing we require a negative polarity to delay or stop this process.



8. Summary

CTE and dark current generation are two critical performance parameters that remain uncertain for the Sandbox CCDs. These characteristics have varied dramatically in the past depending on the quality of the silicon wafers used. The concern is important and most manufacturers have experienced the difficulty. CCDs require the best silicon that can be grown and efforts in this area have been given top priority. Curiously silicon material was not a major concern to CCD performance in the past. Other CCD problems related to process and design difficulties were often dominating. Now that the CCD has matured in these areas silicon material has become the limiting factor for many performance parameters. Unfortunately, no method has been devised to test silicon wafers before CCDs are fabricated. Finding a technique, equally as sensitive as the CCD, is a major challenge for CCD workers. For now we can partially deal with the silicon problem, if it arises, by how the CCD is operated. This paper, for example, has discussed several operational techniques to improve CTE by the way the CCD is cooled and clocked. However, to be effective, the user must perform a series of tests to understand the CTE mechanisms involved (i.e., characterize bulk states).

The silicon problem will likely become more critical in the near future because of four reasons. First, as CCD arrays get larger CTE performance will need to keep pace. Second, dark current will be important because cooling requirements for large arrays will be demanding. Third, 4-inch material, which many CCD manufacturers now use, will eventually become obsolete (3-inch wafers have gone through a similar history). Quality silicon will come from 5-, 6-inch or larger wafer sizes. Four inch manufacturers are already realizing this fact. Many CCD manufacturers in the states are moving to larger CCD lines which will relax the problem. (e.g., the Reticon CCD group is changing over to a 6-inch line). Fourth, new scientific applications will demand improved performance which in turn will push CCD technology and silicon requirements.

Three of the 4096 x 4096 Sandbox CCDs are some of the largest CCDs made today. Currently the CCD is going through a growth period where prototype arrays larger than this have been fabricated. As long as yield figures continue to improve for the CCD larger arrays can be fabricated. An ultra-clean process facility in conjunction with large silicon wafers are key requirements in fabricating large CCDs. For example, the Philips CCD group in the Netherlands has an ideal laboratory to fabricate ultra-large scientific CCD arrays. Their CCD lab is based on a Class-1, 6-inch wafer line. The scientific community may take advantage of this opportunity later this year and fabricate a 9k x 7k, 12-micron pixel CCD, the largest monolithic chip fabricated. Other CCD groups have similar plans. It will be interesting to see just how big the CCD will become in the next few years. Currently it is difficult to predict the theoretical size limitation as things improve.

Competition among CCD manufacturers is currently intense. This has not always been the case. Several years ago scientific CCD production was in jeopardy forcing some good groups out of business. CCD technology has finally matured and has generated an explosion of applications making it now profitable for many groups. However, cost for a custom CCD remains to be a problem for the scientific community in general. The price for a lot run is approximately \$100K to \$140K and continues to rise rapidly. Additional funds are required to perform testing and packaging tasks. Thinning and backside illumination requirements typically doubles the price. The Sandbox method described in this paper attempts to reduce costs by sharing lot expenses among several customers. Cost for Sandbox customers depends on array size and quantity of devices that occupy the wafer.

An exciting area to watch in the future is high speed CCD advancements. This paper briefly addresses some high speed limitations for the CCD. Knowing these problems allows the CCD to be designed and tailored to perform amazing high speed feats. How far the CCD advances in this area depends significantly on computer technology which has advanced at the same pace as the CCD. The Cinema CCD, for example, challenges this market because storage requirements are very demanding. A ten minute movie scene, for example, will generate 900 giga-bytes of information (2 bytes / pixel). A tera-byte computer system is therefore required to run the Cinema CCD. Although building a camera and computer system is technically possible, it remains to be seen if profits can be made on such a camera system.

9. ACKNOWLEDGEMENTS

We acknowledge the many talented CCD researchers, technicians, engineers, scientists and managers that we have worked with over the years. Thanks to Dick Bredthauer, Paul Vu, Charles Chandler, and Ed Jakl at Loral Fairchild for producing many fine CCDs. We appreciate the insight into CCD characteristics from Morley Blouke at SITE. Special thanks to John Vargas at EG&G Reticon for reviewing all Sandbox CCD designs and making numerous corrections. Thanks to Paula Grunthaler and Shouleh Nikzad for many rewarding discussions on MBE backside CCD processing. Thanks goes to Joe Milelli at Loral Fairchild for supplying test CCDs and useful technical conversations. We appreciate the UV QE data supplied in Figure 20b by Craig Staller

of JPL. We thank from Andy Collins for his help in reviewing this long paper.

We are indebted to our Sandbox CCD customers: (1) Cinema - Jim Foster and Paul Jenkins of Advanced Digital Imaging; (2) Advanced Camera - Jim Fanson, Art Vaughan, Mike Devirian, John Truager, Steve Pravdo, and Dave Swenson of JPL; (3) MACH II - Paul Dimotakis and Dan Lang of CIT; (4) Adapt II - Marty Levine and Rick Wilson of JPL and Chris Shelton of Mt. Wilson Observatories; (5) Pluto Flyby - Pat Beauchamp, Chris Stevens, and George Fraschetti of JPL and; (6) Big CIT - Jim McCarthy of CIT and Steve Synnott of JPL.

Research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under several contracts with the National Aeronautics and Space Administration. Reference herein to a specific product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or Jet Propulsion Laboratory, California Institute of Technology.

10. REFERENCES

Many Scientific CCD topics examined in this paper are discussed in considerable depth in a collection of CCD publications written or co-authored by the CCD Advanced Development Group at JPL. These articles and papers can be obtained on request by telephone or letter. Several CCD publications were collected and published in three CCD special issues by Optical Engineering: (1) Charge Coupled Device Characterization, Modeling and Application, Volume 26 Number 8, August 1987; (2) Charge Coupled Device Manufacturer and Application, Volume 26 Number 9, Sept. 1987; (3) Charge Coupled Device and Charge Injection Device Theory and Application, Volume 26 Number 10, Oct. 1987.

- 1). Janesick, J., and Blouke, M., "Past and Present Scientific CCDs," Optics and Photonics News, April, 1995.
- 2). Wizenread, R., "Flat, Thinned Scientific CCDs," SPIE #2198-1212 (March, 1994).
- 3). Pinter, J., and Janesick, J., "CCD Modeling", presented in the 1995 IEEE Workshop on Charge-coupled Devices and Advanced Image Sensors, Dana Point CA., April 20-22.
- 4). Nikzad, S., Hoenk, M., Grunthaner, P., Terhune, R., Gruntahner, F., Wizenread, R., Fattahi, M., and Tseng, H., "Delta-doped CCDs: High QE with Long -term stability at UV visible wavelengths," SPIE, #2198-1212 (March, 1994).
- 5). Blouke, M., Delamere, A., and Womack, G., SPIE Vol. 1447, 142-155, (1991).
- 6). Saks, N., and Brown, D., "The Role of Hydrogen in Interface Trap Creation by Radiation in MOS Devices - a Review," Proceedings of the 2nd Symposium on the Physics and Chemistry of SiO₂ and the Si-SiO₂ Interface, St. Louis, MO, May 18- 21, 1992